PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G06F 19/00

A2

(11) International Publication Number: WO 00/70528

(43) International Publication Date: 23 November 2000 (23.11.00)

(21) International Application Number: PCT/US00/13154

(22) International Filing Date: 12 May 2000 (12.05.00)

(30) Priority Data:

09/310,879	14 May 1999 (14.05.99)	US
09/311,890	14 May 1999 (14.05.99)	US
09/311,996	14 May 1999 (14.05.99)	US
60/134,104	14 May 1999 (14.05.99)	US
Not furnished	26 April 2000 (26.04.00)	US

(71) Applicant: CYTOKINETICS, INC. [US/US]; Suite 2, 280 East Grand Avenue, South San Francisco, CA 94080 (US).

(72) Inventors: SABRY, James, H.; 52 Buena Vista Terrace, San Francisco, CA 94117 (US). ADAMS, Cynthia, L.; 615 Georgia Avenue, Palo Alto, CA 94306 (US). VAISBERG, Eugeni, A.; 647 Pegasus Lane, Foster City, CA 94404 (US). CROMPTON, Anne, M.; 2 Bellair Place, San Francisco, CA 94133 (US). BLUM, Robert, I.; 17 Shoreview Avenue, San Francisco, CA 94121 (US). OESTREICHER, Donald, R.; 904 Old Town Court, Cupertino, CA 95014–4024 (US). SIGAL, Nolan, H.; 941 Berry Avenue, Los Altos, CA 94024 (US).

(74) Agent: LOUIE, Michael, L.; Beyer Weaver & Thomas, LLP, P.O. Box 130, Mountain View, CA 94042–0130 (US).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

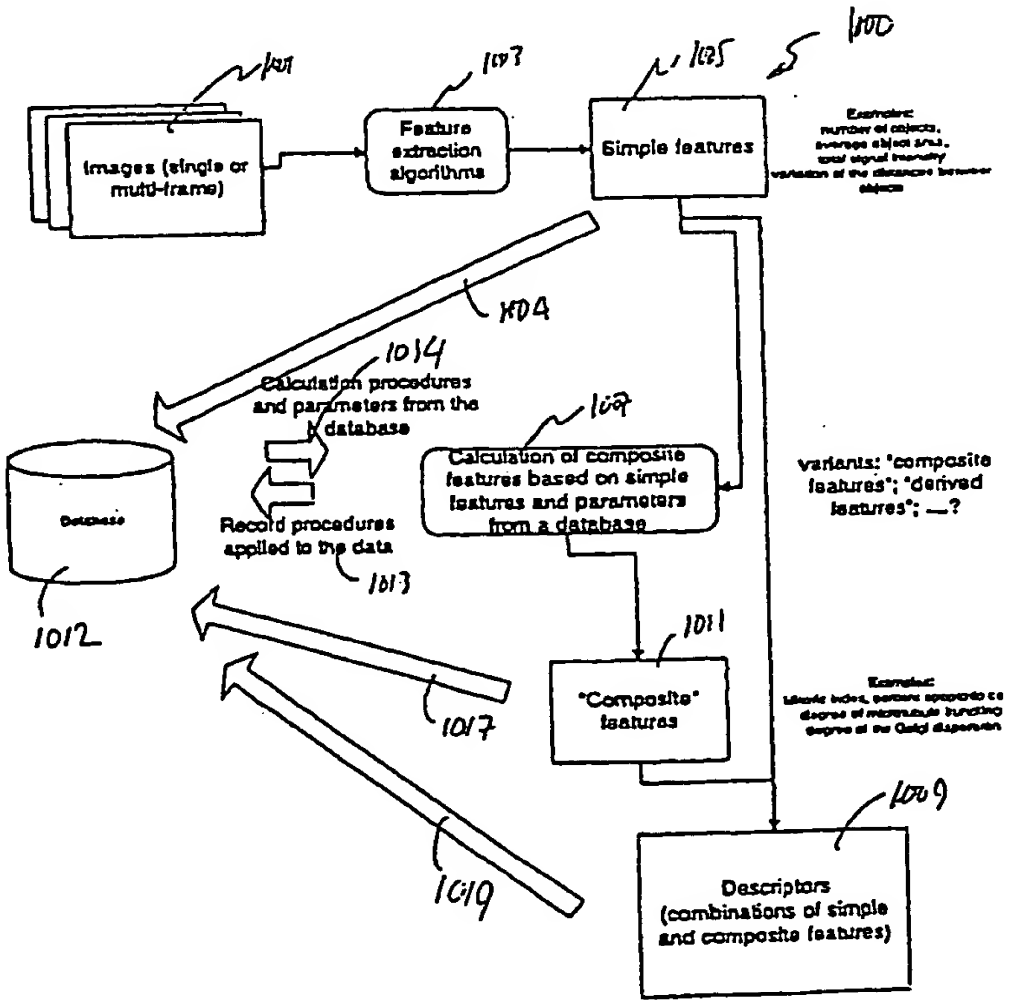
Published

Without international search report and to be republished upon receipt of that report.

(54) Title: METHOD AND APPARATUS FOR PREDICTIVE CELLULAR BIOINFORMATICS

(57) Abstract

Techniques for using information technology in therapeutics or drug discovery. In an exemplary embodiment, techniques for determining information about the properties of substances based upon information about structure of living or non–living cells exposed to substances are provided. A method according to the present invention enables researchers and/or scientists to identify promising candidates in the search for new and better medicines or treatments using, for example, a cellular informatics database. The present invention further teaches a system for acquiring knowledge from cellular information. The system has a database 1012 comprising a database management module ("DBMS"). The system also has a variety of modules, including a population module coupled to the DBMS for categorizing and storing a plurality of features (e.g., cell size, distance between cells, cell population, cell type) from an image acquisition device into the database. The system has a translation module coupled to the DBMS for defining a descriptor from a set of selected features from the plurality of features. In a specific embodiment, the descriptor is for a known or unknown compound, e.g., drug. A prediction module is coupled to the DBMS for selecting one of a plurality of descriptors from known and unknown compounds from the database based upon a selected descriptor from a selected compound. The selected compound may be one that is useful for treatment of human beings or the like.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

PATENT APPLICATION
METHOD AND APPARATUS FOR
PREDICTIVE CELLULAR BIOINFORMATICS

COPYRIGHT NOTICE

BACKGROUND OF THE INVENTION

The present invention provides techniques for information management using a database platform. More particularly, the present invention provides a system including computer code that couples to a database device. The system provides for image capturing of living, dead, or fixed cells or cell fractions used to identify information about substances used on the cells or information about the cells themselves. Accordingly, the present invention can enable researchers and scientists to identify promising candidates in the search for new and better medicines, for example, in drug discovery and development. The principles enumerated herein may, with equal facility, be applied to other applications, including but not limited to use in environmental applications such as determining chemical toxicities and other non-pharmaceutical toxicology uses.

For a long time, researchers in the pharmaceutical field have sought for better ways of searching for substances possessing properties that make them suitable as medicines. In the early days, researchers generally relied upon extracts from plants, dyes, and microbiological extracts for such substances. Examples of such substances include the pain reliever aspirin, the anti-cancer drug paclitaxel (brand name Taxol™), and the heart medication called digoxin. The number of useful medicines has generally been limited.

Purified substances having desirable bio-active properties are also often difficult to discover. Advances in traditional organic chemistry and more recently the rapid chemical synthesis methods often referred to as combinatorial chemistry have increased the number of compounds that researchers test for biological activity. Originally, substances were often initially tested on animals or humans to determine their biological activity. While results from such tests may identify a good drug candidate, they are often time consuming and costly, thus a limited number of substances can be tested. Therefore, pharmaceutical companies have turned to testing their ever-increasing libraries of substances against isolated proteins (drug targets) in biochemical assays that can be carried out at high throughput and low cost. It should be noted that the substances need to be tested in numerous protein tests, each customized for a particular drug target. Therefore, although each protein test may be run at a high-throughput, the design of multiple protein tests can be time-consuming. Substances deemed promising based on results from the protein tests are then tested in lower throughput cellular and animal tests.

There have been some attempts to use image acquisition techniques to screen a large number of substances based upon biological cell information. One such attempt is described in International Application No. WO 98/38490 in the names of Dunlay, et al. Dunlay et al. generally describes a conventional image acquisition system. This conventional system collects and saves images based on certain criteria that are predefined, not on a fixed area of an imaging surface. Additionally, the conventional system has poor lighting design, which makes image processing for multiple cells difficult. Furthermore, the conventional system is not designed for capturing, populating and utilizing a large database design. The conventional system is designed for customized cellular assays, not as a tool for generation of a cellular informatics database. Without such database capabilities the conventional system cannot be used for screening, analyzing, and comparing large quantities of cells from multiple experiments on multiple days in a predictive, efficient and cost effective manner.

What is needed is a rapid assay to assess the activity of compounds against multiple drug targets simultaneously in a cellular context. What is also needed are techniques for finding the effects of substances on cell function based upon searching and analyzing cellular information.

SUMMARY OF THE INVENTION

According to at least one embodiment of the present invention, techniques for determining information about effects of potential substances on cells are provided. In another exemplary embodiment, the present invention provides a novel system including hardware, computer codes, user interfaces, and a database for acquiring, storing and retrieving cellular and substance information. The cells can include living, dead, or fixed cells or fractions of cells. The present invention enables, *inter alia*, researchers and/or scientists to identify promising candidates in the search for new and better medicines or treatments using, for example, a cellular informatics database.

According to the present invention, a computer program for identification and verification of biological properties of substances can include code that causes a sample of a substance to be administered to a cell. The code determines one or more features for two or more cell components, or markers, in the presence of the substance. The code can form one or more descriptors from the features. Descriptors can be formed by combining features of two or more cell components as identified using the markers. The code can then search one or more descriptors obtained from prior administered substances upon cells in order to locate descriptors having a relationship to the descriptors noted for the substance under study. The code predicts properties of the administered substance based upon the properties of the prior administered substances using the relationship between the descriptors. The code can provide for identifying properties of substances based upon effects on cell characteristics. Candidate drug mechanisms of action, potency, specificity, pharmacodynamic, and pharmacokinetic parameters, toxicity, and the like can be used as substance properties.

In a specific embodiment, the present invention provides a system for acquiring knowledge from cellular information. The system has a database comprising a database management module ("DBMS"). The system also has a variety of other modules, including a population module that is coupled to the DBMS and serves to categorize and store a plurality of features (including but not limited to cell size, distance between cells, cell population, as well as sub-cellular features such as organelle location, protein location and sub-cellular constituent location and

movement) from an image acquisition device into the database. The system has a translation module coupled to the DBMS for defining a descriptor from a set of selected features from the plurality of features. In a specific embodiment, the descriptor is for a known or unknown compound, e.g., drug. A prediction module is coupled to the DBMS for selecting one of a plurality of a descriptors from known and unknown compounds from the database based upon a selected descriptor from a selected compound. The selected compound may be one that is useful for treatment of human beings or the like.

In a specific embodiment, the present invention provides a system for populating a database with cellular information. The system includes a cell holder (e.g., multi-well plate, chip, microfluidic assembly, or other cell chamber) comprising a plurality of sites in a spatial orientation. Each of the sites is capable of holding a plurality of cells to be imaged. Note – the light guide is one embodiment, but we don't want to be limited to it.

According to one embodiment, the present system also has an illumination apparatus including a liquid light guide operably coupled to the imaging device for highlighting the plurality of cells in a relatively even spatial manner for image capturing and measurement purposes. Still further, the liquid light guide allows sub-elements (e.g., filter, lamp) of the illumination apparatus to be placed at a remote location to prevent mechanical interference of the cell holder during image capturing. Alternative lighting methodologies may, with equal facility, be implemented.

The system also has an image-capturing device (e.g., charge coupled device camera, translation stage, shutter, microscope, software, shutter control) coupled to a computing device (e.g., computer, network computer, work station, analog computing device, on-board image-processor, and laptop). The image-capturing device is adapted to capture at least one image in at least one of the plurality of sites. One some embodiments, multiple images can be captured, where each image represents a different cell component (or portion). The image-capturing device can be adapted to convert the image into a digital representation, which highlights the feature or features of the one site.

A database storage device (e.g., relational database, object oriented database, mixed object oriented database) includes a database management element. The

database is coupled to the image capturing device. In a specific embodiment, the present system includes modules for feature extraction, generation of descriptions, and data preparation and analysis.

In a specific embodiment, the present invention provides a novel system for determining an effect of a manipulation of a cell using one or more image frames. The system has a plate comprising a plurality of sites in a spatial orientation. Each of the sites is capable of holding a plurality of cells to be imaged. The system also has an image capturing device to capture a plurality of images of at least one site from the plurality of sites. The image capturing device is coupled to the computing device. The system also has an image processing device to combine the plurality of images of at least one site or plurality of sites. The image processing device is operably coupled to the plate. An image processing device is also included. The image processing device can be adapted to form a digitized representation of the plurality of images from the site or plurality of sites. Furthermore, the system has a database storage device comprising a database management element. The database can be adapted to retrieve the descriptor or descriptors of the plurality of features from the computing processing device and storing them in a selected manner.

In a specific embodiment, the present invention provides a system for capturing cellular information. The system also has an image acquisition system comprising a charged coupled device camera adapted to capture an image of a plurality of manipulated cells in various stages of the cell cycle. The stages of the cell cycle are currently understood to include interphase, G0 phase, G1 phase, S phase, G2 phase, M phase, prophase, prometaphase, metaphase, anaphase, and telophase. The principles of the present invention specifically contemplate the application thereof on additional cell cycle stages when and if they are identified.

An optical source is coupled to the image acquisition system for highlighting the plurality of manipulated cells in the various stages of the cell cycle. The illumination apparatus provides for an acquisition of the image of the plurality of manipulated cells. In a specific embodiment, the illumination apparatus has a liquid light guide coupled to a light source at a remote location.

A variety of user interfaces are utile for accessing the several features of the present invention. Those having ordinary skill in the art will appreciate that different user interfaces may be required to support different research scenarios. The

present invention specifically contemplates the utilization of a wide variety of user interfaces.

Numerous benefits are achieved by way of the present invention over conventional techniques. The present invention can provide techniques for predictive cellular bioinformatics that can streamline a number of important decisions made in the drug discovery industry. The present invention can be implemented using off the shelf hardware including databases. In other aspects, the present invention can find useful information about substances as well as cells or portions of cells. Furthermore, the present invention can acquire more then one feature using more than one manipulation. Moreover, the present invention can provide information about a wide variety of cellular information that is not conventionally available. This information includes information about different cell components, e.g., nuclei and Golgi apparatus. Still further, the present invention provides an automated or semi-automated technique for acquiring images and populating a database. The present database can be combined with others such as genomics, and the like. Moreover, the present invention can be implemented to predict, *inter alia*, a mechanism of action, toxicity, target validation, and pre-clinical disease model.

A further understanding of the nature and advantages of the invention herein may be realized by reference to the remaining sections of the specification and the attached drawings.

BRIEF DESCRIPTION OF THE DRAWING

For more complete understanding of the present invention, reference is made to the accompanying Drawing in the following Detailed Description of the Invention. In the drawing:

Fig. 1 is a simplified system diagram according to an embodiment according to the present invention;

Figs. 1A-1B are more detailed diagrams of database systems according to embodiments of the present invention;

Fig. 2 is a simplified block diagram according to an alternative embodiment according to the present invention;

Figs. 3-6 are simplified diagrams of system elements according to embodiments of the present invention;

Figs. 7A-7K illustrate representative block diagrams of simplified process steps in a particular embodiment according to the present invention;

Fig. 8A-8F illustrate representative quantified descriptors of effects of manipulations on images of cells in a particular experiment;

Fig. 9 illustrates example images for different types of morphologies in a particular experiment;

Fig. 10 illustrates a distribution of various morphologies in a cell population responsive to drug concentration in a particular experiment;

Fig. 11 illustrates a graph of quantified features of effects of manipulations on cells in a particular experiment;

Fig. 12 illustrates effects of external agents on cells in a particular experiment;

Fig. 13 illustrates 4 panels for each marker for a plurality of A549 cells in a particular experiment;

Fig. 14 illustrates 4 panels for each marker for a plurality of OVCAR-3 cells in a particular experiment;

Fig. 15 illustrates 4 panels for each marker for a plurality of OVCAR-3 cells at 20x in a particular experiment;

Fig. 16 illustrates 4 panels for each marker for a plurality of OVCAR-3 cells at 40x in a particular experiment;

Fig. 17 illustrates a representative input for a morphometric analysis program in a particular embodiment according to the present invention; and

Figs. 18-19 illustrate examples of the generation of pseudo-sequences and clustering in a particular embodiment according to the present invention.

Fig. 20 is a block diagram for a first research scenario;

Fig. 21 is a block diagram for a second research scenario; and

Fig. 22 is a block diagram for a third research scenario.

Reference numbers refer to the same or equivalent parts of the invention throughout the several figures of the Drawing.

DETAILED DESCRIPTION OF THE INVENTION

According to the present invention, techniques for determining information about manipulated cells or substances based upon living, fixed, or dead cell structures or portions of cells are provided. In an exemplary embodiment, the present invention provides a novel system including computer codes coupled to a database and user interfaces for acquiring, storing and retrieving such information. Other embodiments provide a novel image capturing system for providing digitized representations of live and dead cell structures or the like.

Fig. 1 is a simplified system diagram 10 of a cellular knowledge-based system according to an embodiment of the present invention. This diagram is merely an example and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. The present system 10 includes a variety of elements such as a computing device 13, which is coupled to an image processor 15 and is coupled to a database 21. The image processor receives information from an image capturing device 17, which image processor and image capturing device are collectively referred to as the imaging system herein. The image capturing device obtains information from a plate 19, which includes a plurality of sites for cells. These cells can be biological cells that are living, fixed, dead, cell fractions, cells in a tissue, and the like. The computing device retrieves the information, which has been digitized, from the image processing device and stores such information into the database. A user interface device 11, which can be a personal computer, a work station, a network computer, a personal digital assistant, or the like, is coupled to the computing device.

Fig. 1A is a simplified diagram of a database system 1000 according to an embodiment of the present invention. This diagram is merely an example and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize many other variations, modifications, and alternatives. Database system 1000 includes a variety of techniques for processing images from biological cells, e.g., fixed, living, and dead cells, and cell portions. As shown, images are acquired 1001. These images can be from a single frame or multiple frames. As merely an example, an image processing system may analyze such images. One example of

such an image processing system is described below, but should not be construed as limiting certain claims.

In a specific embodiment, cell samples are manipulated using a compound (e.g., substance, drug). The cell samples are imaged for a simple portion or portions, e.g., manipulated cell substructure, manipulated spatial feature of cell, cell density. Image processing techniques are used to extract 1003 the feature or features from the image or images. The features can be an independent or a dependent set of cell characteristics (which may be predominately visual) including, for example, count, area, perimeter, length, breadth, fiber length, fiber breadth, shape factor, elliptical form factor, inner radius, outer radius, mean radius, equivalent radius, equivalent sphere volume, equivalent prolate volume, equivalent oblate volume, equivalent sphere surface, average intensity, total intensity, optical density, radial dispersion, texture difference, and others. Each of these features corresponds to a similar manipulation by a compound. Each manipulation forms a new set of features, which are identifiable to the compound. Once each set of features has been extracted, the feature set is populated 1004 into a database 1012. Accordingly, the database includes many sets of features, where each set corresponds to a different manipulation for a selected cell. Each set of features corresponding to a manipulation provides a descriptor 1009, which is also stored 1019 in the database. The descriptor is a "finger print" including each feature for the manipulation. Each descriptor may be unique, or may have similarities to other descriptors or may even be the same as other descriptors for known and unknown manipulations.

The present system retrieves features, which we define as simple features herein, and forms composite features 1007 from them. More than one feature can be combined in a variety of different ways to form these composite features. In particular, the composite feature can be any function or combination of a simple feature and other composite features. The function can be algebraic, logical, sinusoidal, logarithmic, linear, hyperbolic, statistical, and the like. Alternatively, more than one simple feature can be combined in a functional manner (e.g., arithmetic, algebraic). As merely an example, the composite feature equals a sum of feature 1 and feature 2, where these features correspond to the same manipulation. Alternatively, the composite feature equals feature 1 divided by feature 2. Alternatively, the composite feature equals feature 1 minus feature 2. Alternatively,

the composite feature equals a constant times feature 1 plus feature 2. Of course, there are many ways that the composite feature can be defined. The present system also stores 1017 these features in the database. The composite features can also be further combined with simple features. Once these features are defined as descriptors, they are stored 1019 in the database.

Fig. 1B is a simplified diagram of a database system engine 2000 according to an embodiment of the present invention. This diagram is merely an example and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize many other variations, modifications, and alternatives. The engine can be implemented into the present database for populating, searching, and predicting compound or cell characteristics. As merely an example, engine 2001 includes an input/output module 2008. The input/output module is used to input and output information from the database. The information includes, among others, a plurality of feature sets, which correspond to many manipulations. Additionally, the information includes descriptors, which each corresponds to a set of features from the manipulation. The database also has a population module, which is used to configure the features based upon an entity relationship, which has been predetermined.

The database engine also has other modules. In particular, the database has a transcription module, which transfers a preselected set of features and creates a descriptor from them. The transcription module can be used to take a known compound, which has features, to transcribe them into a descriptor. Alternatively, the transcription module can be used to take an unknown compound, which has features, to transcribe them into a descriptor. These descriptors are provided into the database for subsequent use. Finally, the database engine has a prediction module, which can be used to potentially predict a property (e.g., mechanism of action) of an unknown compound. Here, the unknown compound is provided with a descriptor, but the property of the compound is unknown. In one embodiment, the prediction module compares a descriptor of an unknown compound with the many descriptors of known compounds, which were in the populated database. Depending upon the matching criteria, the prediction module will attempt to uncover one or more descriptors of known compounds. Once the prediction module finds the descriptors of the known compounds based upon the descriptor for the unknown compound, it identifies a potential property of such unknown compound for analysis and review. Here, it is

believed that certain features of the known compound, which are similar to those features of the unknown compound may uncover a property to the unknown compound. Details of the present software engine are described more fully below.

Fig. 2 is a simplified block diagram 20 of a cellular knowledge-based system according to an alternative embodiment of the present invention. This diagram is merely an example and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. Like reference numerals are used in the present diagram as the previous diagram for easy cross-referencing, but are not intended to be limiting in any manner. The present diagram 20 includes a variety of elements such as a processor 13 or computing device coupled to a database 11. The processor can be used for retrieving and storing information from the database. The system also includes a plurality of system elements, such as a cleaner 23, a dispenser 25, and an image capturing system 27, which are also coupled to the database in some embodiments. These elements can be coupled to each other through a network or the like. As merely an example, the network can be a NetWare™ network from Novell Corporation or an internet network or the Internet but can also be others and any combination thereof. The system also has an output device 31, which can be used to output information from the database, processor, or other system elements. Details of these elements are described more fully below in reference to the Figs.

Figs. 3-5 are simplified drawings of system elements according to embodiments of the present invention. These diagrams are merely examples and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. As merely an example, Fig. 3 is a simplified diagram of a processor or computing device 13. The computing device 13 includes a bus 112 which interconnects major subsystems such as a central processor 114, a system memory 116 (e.g., random access memory), an input/output ("I/O") controller 118, an external device such as a display screen 124 via a display adapter 126, a keyboard 132 and a mouse 146 via an I/O controller 118, a SCSI host adapter (not shown), and a floppy disk drive 136 operative to receive a floppy disk 138.

The computing device has other features. Storage Interface 134 may act as a storage interface to a fixed disk drive 144 or a CD-ROM player 140 operative

to receive a CD-ROM 142. Fixed disk 144 may be a part of computing device or may be separate and accessed through other interface systems. A network interface 148 may provide a direct connection to a remote server via a telephone link or to the Internet. Network interface 148 may also connect to a local area network ("LAN") or other network interconnecting many computer systems. Many other devices or subsystems (not shown) may be connected in a similar manner. Also, it is not necessary for all of the devices shown in Fig. 3 to be present to practice the present invention, as discussed below. The devices and subsystems may be interconnected in different ways from that shown in Fig. 3. The operation of a computer system such as that shown in Fig. 3 is readily known in the art and is not discussed in detail in this application. Computer code to implement the present invention, may be operably disposed or stored in computer-readable storage media such as system memory 116, fixed disk 144, CD-ROM 140, or floppy disk 138. The computer code can be organized in terms of processes or modules, depending upon the application. That is, the computer code can include a prediction module, a translation, module, or other modules to carryout the functionality described herein, as well as others.

Figs. 4 and 5 are simplified diagrams of an imaging system 200 according to an embodiment of the present invention. As shown, the imaging system 200 includes a variety of features such as housing 203, which holds a stage assembly 204. The stage assembly includes an x-stage movement element 206, which is along an x-direction, and a y-stage movement element 207, which is along a y-direction. The imaging system also includes a z-direction movement element, which is perpendicular to the x-y plane. The z-direction movement motor can be attached to the stage, or to the objective nosepiece by way of the microscope housing, or as an external motor between the objective and the microscope housing. The stage can align in any one of the directions to an accuracy of one micron and less, or one-half micron and less, or one-quarter micron and less, depending upon the embodiment.

The stage holds a plate 202 or cell holder, which houses one of a plurality of samples. The plate includes a spatial array 209 of process sites. Each of the process sites can include a plurality of cells and solutions depending upon the embodiment. Each of the sites can carry a sufficient amount of solution to prevent substantial evaporation of the sample during processing in some embodiments. In embodiments for large scale analysis, the plate includes at least 96 sites, or more than

or equal to 384 sites, or more than or equal to 1,536 sites. The plate bottom is transparent and thin, which allows light to pass through the sample. Additionally, the plate is made of a suitable chemical resistant material. As merely an example, the plate can be either a 96, or 384, or 1536 or other formats from places such as Becton Dickinson of Franklin Lakes, NJ, or Corning Science Products of Corning, NY. In a preferred embodiment, the plate is a Corning Costar black-walled 96 well plate catalog #3904 from Corning Science Products of Corning, NY, but should not be limited to these in some applications, but can be others.

Also shown is the condenser for the microscope 201, which can be used to collect phase, DIC, or bright field images of the cells. Images resulting from the illumination of the samples to fluorescence, phase, DIC, or bright field techniques are collected using an image capturing device 208, which captures an image or images of cells from the plate. In a specific embodiment, the microscope is an inverted configuration with the objectives on the bottom of the plate and the condenser disposed overlying an upper surface of the sites, while the image capturing device underlies the sites. Images captured by the imaging device, whether analogue or digital, are viewed by a monitor or other devices. The image capturing device can be any camera assembly such as a charge coupled device camera, which is known as a CCD camera, or other high resolution camera capable of capturing images from the sites. In a specific embodiment, the camera is an interline CCD camera which does not require an external shutter.

In a specific embodiment, the present imaging system can be any suitable unit that is flexible for automated image collection using multi-well plastic plates. The imaging system also should be adapted to collect high-resolution images of cells on plastic or glass plates, cell growth chambers, or coverslips. The system also can be used for imaging multiple cell markers in multiple imaging conditions. To accomplish this, the microscope system has a variety of elements such as a light source, a motorized excitation filter wheel and shutter, x-y-z-motorized stage, excitation and emission filters, Fluor phase and DIC objectives, motorized objective nosepiece, dichroic filters, motorized dichroic filter cubes, phase and DIC rings and prisms, CCD camera, and software control. As merely an example, the present imaging system can have components such as those listed in the Table below.

DESCRIPTION	MAKER	MODEL
Microscope	Zeiss	100M
(x-y) motorized stage	Prior	
Xenon lamp	Sutter	Lambda
Filter wheel	Sutter	Lambda-10
Microtitre Plate holder	Prior	500-H223R
Isolation Table	Kinetic Systems	9101-24-85
Objective Spacers	Polytec PI	P-721.90
Camera	Hamamatsu	C47-95
Computer	IBM	IntelliStation
Software	Metamorph	v.4
Objectives	Zeiss	Achroplan 10x/0.25 LD-Achroplan 20x/0.4 LD-Achroplan 40x/0.6

Table: Image Acquisition System Elements

In a specific embodiment, the present system has the following capabilities, which are not intended to be limiting.

Image acquisition

1) Ability to automatically acquire multi-wavelength images from multiple sites on one multi-well plate, to sequentially name image files, and to log any imaging parameter information with image files.

2) Ability to link images with a larger database/spreadsheet of information.

3) Ability to automatically collect multiple plates by interfacing the imaging system with a robotic arm.

X-Y control

1) Ability to place 96, 384, or 1536 well plates onto microscope stage and move to each well sequentially.

2) Ability to return to each well and collect another round of images (multi-site time-lapse) or ability to collect rapid time-lapse information at each well (time-lapse of many wells).

3) Ability to collect a low magnification image, automatically determine features which may be of interest, automatically change the objective to a higher magnification, and collect high magnification images of a fixed number of those identified cells in the sample.

4) Ability to collect multiple frames in each site.

Z control

1. Ability to auto-focus with substantially minimal damage to biological specimen or fluorophore.

2. Ability to auto-focus rapidly.

The present embodiment of the imaging system is shown by way of Figs. 5A and 5B. These diagrams are merely examples and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. The present imaging system 40 includes a variety of elements such as a microscope 41, which is preferably an epi-fluorescent microscope, but can be confocal, multiphoton, or hybrid types. The microscope includes elements 41A, the motorized Z-axis; 41B, the motorized dichroic filter cube holder; and 41C, the motorized objective nosepiece. In one embodiment, the microscope is a Model 100M made by Zeiss. The microscope communicates to computer 51 through control lines 73, 75, and 76. The imaging system also has camera 50 coupled to controller 50A and computing device 51, which oversees and controls operations of the elements of the imaging system.

The present microscope includes drivers for spatially moving a stage in two dimensions, including an x-direction, a y-direction, and moving the objective nosepiece in a z-direction in a Cartesian coordinate system. The z-direction movement is provided using a fast z-motor, which can make z-direction adjustments within a predetermined time. The z-direction movement generally provides for focussing of the sample to the camera. The focussing occurs within the predetermined time of preferably ten seconds and less, or five seconds and less, or one

second and less, depending upon the embodiment. As merely an example, the z-motor or positioner can be a model PIFOC objective nanopositioner made by a company called Physik Instrumente of Waldbronn, Germany, but also can be others. The z-motor couples to computer 51 through line 63, which may also include a controller. Depending upon the embodiment, a second z-motor 41A connected to the computer 51 by line 73 may be used to keep the z-motor 42 in the center of its travel. Alternatively, in other embodiments the stage could be provided with a z-motor allowing for movement of the stage in the z-direction.

The present stage also includes an x-y stage 43. The x-y stage moves plate 59, e.g., 96 site, 384 site, 1536 site. The x-y stage moves plate in an x-y spatial manner. The stage has an accuracy or repeatability of about 1 micron and less, or about 2 microns and less. The stage can move in a continuous manner or a stepped manner. The stage also can move up to 30 mm/sec. or faster. The stage also can move 1 mm/sec. and less, depending upon the embodiment. The stage can also step 0.1 micron and less or 1 micron and less, as well as other spatial dimensions. The stage can be one such as a Proscan Series made by Prior Scientific of Rockland, MA but can also be others. The stage is controlled via control line 61 through controller 43A, which couples to computer 51 through control line 65.

The stage includes plate holder 44. The plate holder can hold a single plate. In other embodiments, plate holder can also hold multiple plates. The plate holder can use mechanical, electrical, fluid, vacuum and other means for holding the plate or plates. The plate holder also is sufficiently stable for securing the plate. As merely an example, the plate holder is a Model 500-H223R made by Prior Scientific of Rockland, MA . In some embodiments, the plate holder may need adjustment in the z-direction to provide for a desirable focus of a sample on a plate. In these embodiments, the plate holder is supported by spacers 45 or a plurality of stage pins, which mechanically elevate the plate holder in the z-direction. These pins are generally made of a suitable material for supporting such plate holder and also are sufficiently resistant to chemicals and the like.

In some embodiments, the entire imaging system is placed on an isolation table 49. The isolation table is disposed between the microscope and support structure. The isolation table is designed to prevent excessive vibration of the plate. The isolation table is made of a suitable material such as steel and honeycomb but can

be others. The table has a thickness of about 8 inches or preferably less than about 24 inches. In one embodiment, the table is Model 9101-24-85 made by Kinetic Systems of Boston, MA.

The imaging system also has a lamp or illumination assembly 62. The lamp assembly provides for a light source (See reference letter B) to a plurality of elements in the imaging system. For easy reading, the light path is defined by the doted lines, which are not intended to be limiting. The lamp assembly has a variety of elements such as a Xenon lamp 46. The Xenon lamp provides light at about 320 to 700 nanometers (Prefocused). The Xenon lamp is 175 or 300 Watts. As merely an example, the lamp can be a Lambda Model made by Sutter Instrument Company of Novato, CA.

Referring to Fig. 5B, the lamp assembly also has a cold mirror 58, an excitation filter wheel 48, excitation filter(s) 55, and an excitation light shutter 57. As shown, light is derived from the Xenon lamp, reflects off of the cold mirror 58, traverses through the excitation filter or filters 55, and is controlled by the excitation light shutter 57. The lamp assembly has filter wheel 48, which houses one of a plurality of filters, including excitation filters. The shutter and filter wheel are controlled via control lines 67, which are coupled to a computer 51 or other type of computing device. The control lines 67 are coupled through controller 57A (for element 57) and controller 48A (for element 48) via control line 69 to computer 51.

Preferably, light traverses from the lamp assembly through a light guide 47 to illuminate features within the plate. The light guide is suitably selected to have a flexible member, which can be used to place lamp source at a remote location away from the imaging device. The flexible member substantially keeps any vibration from the lamp assembly away from the imaging device. In some embodiments, the member is at least 1 foot away from the imaging device. The light guide is a guide, which is a flexible hose-type sleeve. The sleeve is filled with a liquid such as an aqueous solution containing chloride or phosphate. A thin layer may be formed on the inside of the sleeve. The layer can be a containing tetrafluoroethylene and mexafluoropropylene, or containing tetrafluoroethylene and perfluoromethyl vinyl ether, or tetrafluoroethylene and perfluoropropyl vinyl ether. An example of such a light guide is described in International Application No. WO/98/38537 filed February 29, 1997, and assigned to NATH, Gunther. The liquid

light guide has less than about 30% transmission loss of the light at a remote location such as the imaging system.

Light is derived from the lamp assembly and directs off of filter 56, which directs the light upward. Filter 56 can be a dichroic and emission filter, as well as others. The light traverses through microscope nosepiece 41C, and traverses through objective spacers 54. An objective 53 magnifies the light toward a predetermined point on the plate 59. The objective can be, for example, made by Zeiss of Jena, Germany, as well as other companies. The objective can be one of a plurality including 1X, 10X, 20X, 40X, and others, depending upon the application. Magnification can be further expanded or contracted by intermediate optics between the objective and the camera. Selection of filter or filters is controlled by computer 51 via control line 75.

The camera 50 captures an image of cells from plate 59. The image is obtained from light scattering off of cells or portions of cells in the plate through objective 53, through objective spacers, through filters 56, which are captured at camera 50. In this preferred embodiment, the camera is a digital camera, but can be an analogue camera. The digital camera is a CCD camera, which has 1280 by 1024 pixels, or more or less. The pixels can be 6.7 microns in dimension or more or less. The camera preferably is substantially free from an external shutter to quickly capture a plurality of images of cells from the plate. The camera is controlled via control line 71 through controller 50A, which connects to computer 51 through control line 70. The present invention can also include other types of image acquisition devices selected from at least an epifluorescence, a confocal, a total-internal reflection, a phase, a Hoffman, a bright field, a dark field, a differential interference contrast, an interference reflection, or multi-photon illumination device.

The present imaging system stores images on a high density memory device 60. The high density memory device is preferably optical, but can also be magnetic. The high density memory device can be any suitable unit that is capable of storing a plurality of images from a plurality of sites in the plate. The memory device can be a compact disk, which would generally use a compact disk burner or the like. Depending upon the embodiment, the high density memory device is used to archive the images that are captured from the camera in the imaging system. Further details

of the imaging system can be found throughout the present specification, and more particularly below.

As merely an example, the present invention can be implemented using the following sequence of steps, which have been described in a journal entry form. Here, images are opened and objects are identified based on a background value that has been edited in starting image acquisition. Information is maintained in a spreadsheet or other database format, which has the following information for each object:

Image Name	Image Plane	Image Date and Time
Elapsed Time	Object #	Total area
Pixel area	Area	Hole area
Relative hole area	Standard area count	Perimeter
Length	Breadth	Fiber length
Fiber breadth	Shape factor	Ell. form factor
Inner radius	Outer radius	Mean radius
Average gray value	Total gray value	Optical density
Radial dispersion	Texture Difference Moment	EFA Harmonic 2, Semi-Major Axis
EFA Harmonic 2, Semi-Minor Axis	EFA Harmonic 2, Semi-Major Axis Angle	EFA Harmonic 2, Ellipse Area
EFA Harmonic 2, Axial Ratio	EFA Harmonic 3, Semi-Minor Axis	

After computations are done, the log file is saved. In particular, the file is saved in an appropriate place with an appropriate name.

In a specific embodiment, the present invention provides the following detailed example of journal entries, which should not limit the scope of the invention.

Set Up Sequential File Names	Interactive: user sets up prefix name and image storage directory
Open Data Log	Opens a DDE (Excel) File
Annotate Log File	Interactive: experimental information that will go into the first line of the log file of stage positions
Stage (Go to Origin)	Origin is set as the center of well A1
Stage (Move to Absolute Position)	Offset to upper left hand corner of well (1410, 1621)
Stage (Log Position)	
Stage (Scan Wells)	User picks wells to scan: runs 3x3 image collection.jnl.

3X3 IMAGE COLLECTION.jnl

Stage (Scan)	Takes 9 images of well, -1600 motor steps apart from left to right 3 columns and 3 rows, runs FOCUS, COLLECT IMAGE, SAVE SEQUENTIAL FILE NAME.JNL.

5

FOCUS, COLLECT IMAGE, SAVE SEQUENTIAL FILE NAME.jnl.

Stage (Log Position)	Logs stage position of each image
ADC – Focus	Opens up the manual focusing window with whatever focus time is current set
Show Message and Wait	Interactive: user hits enter to continue when done focusing

ADC-Acquire from Digital Camera	Takes Hoechst image
Save Using Sequential File Names	
Close	Closes image window

START IMAGE ANALYSIS.jnl

Low Pass	3x3 convolution of already opened image
Low Pass	3x3
Show Region Statistics	Interactive: Show entire image statistics. Calculate background subtraction value for step 4. by: INTENSITY Average + INTENSITY Std. Dev.
Arithmetic	Interactive: User inputs subtraction value from 3. into the constant Value field
Threshold image	Creates threshold 1 unit above 0 to 4096
Integrated Morphometry – Load State	Loads Start Image Analysis.ima Classifier 100 < area < 200000
Integrated Morphometry – Measure	Interactive: Shows area summary information about all objects. The average number is used as the Standard Area in 8.
Object Standards - Set Object Standards	Interactive: User inputs average area value from 7. into Standard Area box to be used by automated IMA for all images

IMA OBJECTS.jnl

Low Pass	3x3 convolution
Low Pass	3x3 convolution
Arithmetic	This background subtraction value needs to be manually entered into this journal from the value determined in START IMAGE ANALYSIS.jnl step 3
Threshold Image	1 unit above 0
Integrated Morphometry – Load State	Hoecsht.IMA Classifier 200 < area < 200000
Integrated Morphometry – Measure	Measures statistical info for all objects
Run Journal	Runs log obj and sum data.jnl

Log obj and sum data.jnl

Integrated Morphometry – Log Data	Logs object data into Sheet 1
Integrated Morphometry – Log Data	Log summary data into Sheet 2

5

COLLECT AUTOMATED IMA DATA INONE SPREADSHEET.jnl

Run Journal	Runs OPEN OBJECT LOG DDE FILE.JNL
Loop for all Images in a Directory	Loops IMA OBJECTS.jnl
Close Summary Log	
Close Object Log	User must manually save Excel spreadsheet

OPEN OBJECT LOG DDE FILE.jnl

Open Object Log	Opens a DDE object log into sheet 1 of an Excel spreadsheet
Open Summary Log	Opens a summary log into sheet 2

COLLECT AUTOMATED IMA DATA INONE SPREADSHEET 16 BIT IMAGES.jnl

Arithmetic	Interactive: Opens Arithmetic window for user to input background subtraction level from START IMAGE ANALYSIS.jnl step 3
Run Journal	Runs OPEN OBJECT LOG DDE FILE.JNL
Loop for all Images in a Directory	Interactive: Runs IMA OBJECTS 16 bit.jnl. User picks directory from which to choose.

5

IMA OBJECTS 16bit.jnl

Low Pass	3x3 convolution
Low Pass	3x3 convolution
Copy to 8-bit Image	No autoscale, to new untitled image
Save Using Sequential File Name	Saves 8bit image using previously defined Sequential File names.
Arithmetic	This background subtraction value needs to be manually entered into this journal from the value determined in START IMAGE ANALYSIS16 TO 8 BIT.jnl step 5
Threshold Image	1 unit above 0 to 255

Integrated Morphometry – Load State	Hoecsht.IMA Classifier 200 < area < 200000
Integrated Morphometry – Measure	Measures statistical info for all objects
Run Journal	Runs log obj and sum data.jnl

START IMAGE ANALYSIS 16 to 8 BIT.jnl

Copy to 8-bit Image	No autoscale, to new untitled image
Close	Closes 16 bit image
Low Pass	3x3 convolution
Low Pass	3x3 convolution
Show Region Statistics	Interactive: Show entire image statistics. Calculate background subtraction value for step 6. by: INTENSITY Average + INTENSITY Std. Dev.
Arithmetic	Interactive: User inputs subtraction value from 5. into the constant Value field
Threshold image	Creates threshold by 1 unit above 0 to 255
Integrated Morphometry – Load State	Loads Start Image Analysis.ima Classifier 100 < area < 200000
Integrated Morphometry – Measure	Interactive: Shows area summary information about all objects. The average number is used as the Standard Area in 10.
Object Standards - Set Object Standards	Interactive: User inputs average area value from 9. into Standard Area box to be used by automated IMA for all images

IMA OBJECTS WITH NEW LOG FILE.jnl

Run Journal	OPEN OBJECT LOG DDE FILE.JNL
Run Journal	IMA OBJECTS.jnl
Close Summary Log	
Close Object Log	User must manually save every Excel spreadsheet generated.

INTERACTIVE IMA OBJECTS.jnl

Threshold Image	User manually sets threshold
Integrated Morphometry – Load State	Hoechst.IMA Classifier 200 < area < 200000
Integrated Morphometry – Measure	Objects
Integrated Morphometry – Log Data	Into open object.log file

5

COLLECT INTERACTIVE IMA DATA.jnl

Close Object Lo g	
Open Object Log	Interactive
Annotate Log File	Interactive: experimental information that will go into the first line of the object log file
Loop for all Images in Directory	Runs INTERACTIVE IMA OBJECTS.jnl

CHANGE FILTER, COLLECT IMAGE, SAVE SEQUENTIAL FILE NAME.jnl

Stage (Log Position)	
ADC-Focus	

Show Message and Wait	Interactive – user presses Enter when done focusing
ADC – Acquire from Digital Camera	Hoechst
Save Using Sequential File Name	
Close	Close open image

COLLECT HOECHST AND FITC.jnl

Run Journal	FOCUS, COLLECT IMAGE, SAVE SEQUENTIAL FILE NAME.JNL
Run Journal	CHANGE FILTER, COLLECT IMAGE, SAVE SEQUENTIAL FILE NAME.jnl

3X3 IMAGE COLLECTION HOECHST FITC.jnl

Stage (Scan)	COLLECT HOECHST AND FITC.jnl

5

AUTOMATED 3X3 IMAGE COLLECTION HOECHST FITC.jnl

Set Up Sequential File Names	Interactive: user sets up prefix name and image storage directory
Open Data Log	Excel DDL files
Annotate Log File	Interactive: experimental information that will go into the first line of the log file of stage positions
Stage (Go to Origin)	Origin is set as the center of well A1
Stage (Move to Absolute Position)	Offset to upper left hand corner of well (1410, 1621)

Stage (Log Position)	
Stage (Scan Wells)	Interactive: user picks wells to scan: runs 3X3 IMAGE COLLECTION HOECHST FITC.jnl

AUTOMATED IMAGE COLLECTION.jnl

Set Up Sequential File Names	Interactive: user sets up prefix name and image storage directory
Open Data Log	Opens a DDE (Excel) File
Annotate Log File	Interactive: experimental information that will go into the first line of the log file of stage positions
Stage (Go to Origin)	Origin is set as the center of well A1
Stage (Log Position)	
Stage (Scan Wells)	Interactive: user picks wells to scan: runs FOCUS, COLLECT IMAGE, SAVE SEQUENTIAL FILE NAME.JNL. Well to well travel = (-9035, -9035)

STARTUP.jnl

Install and Configure Devices	Open Stage Meta Devices
Set Live Video Channel	

Preferences	Measure Objects: Draw failed classifier objects, Exclude objects that touch the edge of the image, Enable Elliptical Fourier Parameters, **turn off** Warn users when measurement data will be erased Image Saving: Save Tiff/stk using LZW compression Image Windows: Use transparent thresholds.
Configure Default Paths	C:\Metamorph Data C:\Metamorph Data\Commmon Settings
Load Journal Taskbar	Common.JTB

Nested Journals

Automated 3x3 Image Collection

Loop 3x3 image collection

Loop focus, collect image, save sequential file name

Automated 3x3 image collection Hoechst FITC

Loop 3x3 image collection Hoechst FITC

loop Collect Hoechst and FITC

focus, collect image, save sequential file name

change filter, collect image, save sequential file name

Automated image collection

Loop focus, collect image, save sequential file name

Collect automated IMA data in one Spreadsheet

Open object log DDE file

Loop IMA objects

Log obj and sum data

Collect automated IMA data in one spreadsheet 16 bit images

Open object log DDE file

Loop IMA objects 16 bit

Log obj and sum data

Although the above has been generally described in terms of a specific user interface and software code, other user interfaces and code can also be used. One of ordinary skill in the art would recognize many other variations, alternatives, and modifications.

Fig. 6 is a simplified diagram 600 of a cleaning and dispensing system according to an embodiment of the present invention. This system 600 includes a variety of elements such as a dispensing head 609, which is coupled to a plurality of pipettes 601. The pipettes input and output fluids or solutions from plate 603. The plate has a plurality of sites, each of which can be used to input cells or a combination of cells and solution. The system also has elements to house solutions 605, which are used to manipulate cell samples in the plate. The dispensing head is supported through a support member 607, which is sufficiently rigid to allow for movement of the head. The dispenser is coupled to the present system in a mechanical and electrical manner, which provides for a fully integrated system for providing cell samples to the imaging system according to the present invention.

Fig. 7A illustrates a representative block flow diagram of simplified process steps of a method for determining properties of a manipulation based upon effects of the manipulation on one or more portions of one or more cells in a particular embodiment according to the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. In step 700, one or more samples of cells can be provided. These cells can be live, dead, or fixed cells, or cell fractions. The cells also can be in one of many cell cycle stages, including G0, G1, S, G2 or M phase, M phase including the following cell cycle stages: interphase, prophase, prometaphase, metaphase, anaphase, and telophase.

Cell components tracked in presently preferable embodiments can include proteins, protein modifications, genetically manipulated proteins, exogenous proteins, enzymatic activities, nucleic acids, lipids, carbohydrates, organic and inorganic ion concentrations, sub-cellular structures, organelles, plasma membrane, adhesion complex, ion channels, ion pumps, integral membrane proteins, cell surface receptors, G-protein coupled receptors, tyrosine kinase receptors, nuclear membrane receptors, ECM binding complexes, endocytotic machinery, exocytotic machinery, lysosomes, peroxisomes, vacuoles, mitochondria, Golgi apparatus, cytoskeletal filament network, endoplasmic reticulum, nuclear membrane, proteosome apparatus, chromatin, nucleolus, cytoplasm, cytoplasmic signaling apparatus, microbe specializations and plant specializations.

The following table illustrates some markers and cell components commonly used by embodiments according to the present invention. Other markers can be used in various embodiments without departing from the scope of the invention.

Cell component	Marker	Disease State
Plasma membrane (including overall cell shape)	Carbocyanine dyes Phosphatidylserine Various lipids Glycoproteins	Apoptosis-Cancer Apoptosis-Neural degenerative Ds
Adhesion complexes	Cadherins Integrins Occludin Gap junction ERM proteins CAMs Catenins Desmosomes	Thrombosis Metastasis Wound healing Inflammatory Ds Dermatologic Ds
Ion Channels and Pumps	Na/K Atpase Calcium channels Serotonin reuptake pump CFTR	Cystic fibrosis Depression Congestive Heart Failure Epilepsy

G coupled receptors	β adrenergic receptor Angiotensin receptor	Hypertension Heart Failure Angina
Tyrosine kinase receptors	PDGF receptor FGF receptor IGF receptor	Cancer Wound healing Angiogenesis Cerebrovascular Ds
ECM binding complexes	Dystroglycan Syndecan	Muscular Dystrophy
Endocytotic machinery	Clathrin Adaptor proteins COPs Presenilins Dynamin	Alzheimer's Ds
Exocytotic machinery	SNAREs Vesicles	Epilepsy Tetanus Systemic Inflammation Allergic Reactions
Lysosomes	Acid phosphatase Transferrin	Viral diseases
Peroxisomes/Vacuoles		Neural degenerative Ds
Mitochondria	Caspases Apoptosis inducing factor F1 ATPase Fluorescein Cyclo-oxygenase	Apoptosis Neural degenerative Ds Mitochondrial Cytopathies Inflammatory Ds
Golgi Apparatus	Lens Culinaris DiOC6 carbocyanine dye COPs	

Cytoskeletal Filament Networks	Microtubules Actin Intermediate Filaments Kinesin, dynein, myosin Microtubule associated proteins Actin binding proteins Rac/Rho Keratins	Cancer Neural degenerative Ds Inflammatory Ds Cardiovascular Ds Skin Ds
Endoplasmic Reticulum	SNARE PDI Ribosomes	Neural degenerative Ds
Nuclear Membrane	Lamins Nuclear Pore Complex	Cancer
Proteosome Apparatus	Ubiquityl transferases	Cancer
Chromatin	DNA Histone proteins Histone deacetylases Telomerases	Cancer Aging
Nucleolus	Phase markers	
Cytoplasm	Intermediary Metabolic Enzymes BRCA1	Cancer
Cytoplasmic Signaling Apparatus	Calcium Camp PKC pH	Cardiovascular Ds Migraine Apoptosis Cancer
Microbe Specializations	Flagella Cilia Cell Wall components: Chitin synthase	Infectious Ds

Plant specializations	Choloroplast Cell Wall components	Crop Protection

Then, in a step 702, one or more samples of the manipulation can be provided to the cells. Manipulations can comprise one or any combination of chemical, biological, mechanical, thermal, electromagnetic, gravitational, nuclear, or temporal factors, for example. For example, manipulations could include exposure to chemical compounds, including compounds of known biological activity such as therapeutics or drugs, or also compounds of unknown biological activity. Or exposure to biologics that may or may not be used as drugs such as hormones, growth factors, antibodies, or extracellular matrix components. Or exposure to biologics such as infective materials such as viruses that may be naturally occurring viruses or viruses engineered to express exogenous genes at various levels. Bioengineered viruses are one example of manipulations via gene transfer. Other means of gene transfer are well known in the art and include but are not limited to electroporation, calcium phosphate precipitation, and lipid-based transfection. Manipulations could also include delivery of antisense polynucleotides by similar means as gene transfection. Other genetic manipulations include gene knock-outs or gene mutations. Manipulations also could include cell fusion. Physical manipulations could include exposing cells to shear stress under different rates of fluid flow, exposure of cells to different temperatures, exposure of cells to vacuum or positive pressure, or exposure of cells to sonication. Manipulations could also include applying centrifugal force. Manipulations could also include changes in gravitational force, including sub-gravitation (the preferred embodiment in outer space). Manipulations could include application of a constant or pulsed electrical current. Manipulations could also include irradiation. Manipulations could also include photobleaching which in some embodiments may include prior addition of a substance that would specifically mark areas to be photobleached by subsequent light exposure. In addition, these types of manipulations may be varied as to time of exposure, or cells could be subjected to multiple manipulations in various combinations and orders of addition. Of course, the type of manipulation used depends upon the application.

Then, in a step 704, one or more descriptors of a state in the portions of the cells in the presence of the manipulation can be determined using the images

collected on the imaging system. Descriptors can comprise scalar or vector values, representing quantities such as area, perimeter, dimensions, intensity, gray level, aspect ratios, and the like. Other types of descriptors include, but are not limited to, one or any combination of characteristics such as a cell count, an area, a perimeter, a length, a breadth, a fiber length, a fiber breadth, a shape factor, a elliptical form factor, an inner radius, an outer radius, a mean radius, an equivalent radius, an equivalent sphere volume, an equivalent prolate volume, an equivalent oblate volume, an equivalent sphere surface area, an average intensity, a total intensity, and an optical density. These descriptors can be average or standard deviation values, or frequency statistics from the descriptors collected across a population of cells. These descriptors can be further reduced using other methods such as principal component analysis and the like. In some embodiments, the descriptors include features from different cell portions or cell types. That is, a first feature can be from a nuclei and a second feature is from another cell structure such as Golgi apparatus, mitochondria, spacing between cell structures or cells themselves, as well as many others.

A presently preferable embodiment uses descriptors selected from the following table. Other descriptors can also be used without departing from the scope of the invention.

Name of Parameter	Explanation/Comments
Count	Number of objects
Area	
Perimeter	
Length	X axis
Width	Y axis
Shape Factor	Measure of roundness of an object
Height	Z axis
Radius	
Distribution of Brightness	
Radius of Dispersion	Measure of how dispersed the marker is from its centroid
Centroid location	x-y position of center of mass
Number of holes in closed objects	Derivatives of this measurement might include, for

	example, Euler number (= number of objects - number of holes)
Elliptical Fourier Analysis (EFA)	Multiple frequencies that describe the shape of a closed object
Wavelet Analysis	As in EFA, but using wavelet transform
Interobject Orientation	Polar Coordinate analysis of relative location
Distribution Interobject Distances	Including statistical characteristics
Spectral Output	Measures the wavelength spectrum of the reporter dye. Includes FRET
Optical density	Absorbance of light
Phase density	Phase shifting of light
Reflection interference	Measure of the distance of the cell membrane from the surface of the substrate
1,2 and 3 dimensional Fourier Analysis	Spatial frequency analysis of non closed objects
1,2 and 3 dimensional Wavelet Analysis	Spatial frequency analysis of non closed objects
Eccentricity	The eccentricity of the ellipse that has the same second moments as the region. A measure of object elongation.
Long axis/Short Axis Length	Another measure of object elongation.
Convex perimeter	Perimeter of the smallest convex polygon surrounding an object
Convex area	Area of the smallest convex polygon surrounding an object
Solidity	Ratio of polygon bounding box area to object area.
Extent	proportion of pixels in the bounding box that are also in the region
Granularity	
Pattern matching	Significance of similarity to reference pattern
Volume measurements	As above, but adding a z axis

Then, in a step 705, a database of cell information can be provided. Next, in a step 706, a plurality of descriptors can be searched from a database of cell information in order to locate descriptors based upon one of the descriptors of the manipulation. Then, in a step 708, properties of the manipulation are predicted based upon the properties of the located descriptors. Properties can comprise toxicity, specificity against a subset of tumors, mechanisms of chemical activity, mechanisms of biological activity, structure, adverse biological effects, biological pathways, clinical effects, cellular availability, pharmacological availability, pharmacodynamic properties, clinical uses and indications, pharmacological properties, such as absorption, excretion, distribution, metabolism and the like.

In a particular embodiment, step 706 comprises determining matching descriptors in the database corresponding to a prior administration of the manipulation to the descriptors of the present administration of the manipulation. In a particular embodiment according to the present invention, combinations of measurements of scalar values can provide predictive information. A database can be provided having one or more "cellular fingerprints" comprised of descriptors of cell-substance interactions of drugs having known mechanisms of action with cells. Such descriptors can be analyzed, classified, and compared using a plurality of techniques, such as statistical classification and clustering, heuristic classification techniques, a technique of creating "phylogenetic trees" based on various distance measures between descriptors from various drugs. In this embodiment, numeric values for the descriptors can be used by comparison techniques. A phylogenetic tree can be created that illustrates a statistical significance of the similarity between descriptors for the drugs in the database. Because the drugs used to build the initial database are of known mechanism, it can be determined whether a particular scalar value in a descriptor is statistically predictive. Finally, a compound descriptor with no known mechanism of action can be queried against the database and be statistically compared and classified among the drugs in the database that the compound most resembles.

In a particular embodiment, relationships between measured morphological properties of images and physiological conditions can be determined. Relationships can include, for example, treatment of different cell lines with chemical compounds, or comparing cells from a patient with control cells, and the like. In a presently preferable embodiment, comparisons can be performed on acquired image

features. Some embodiments can comprise statistical and neural network - based approaches to perform comparisons of various features. The foregoing is provided as merely an example, and is not intended to limit the scope of the present invention. Other techniques can be included for different types of data.

In some embodiments, classification, clustering and other types of predictive data analysis can be performed on features extracted from cell images. In a presently preferable embodiment, statistical procedures for comparisons, classification and clustering are performed on data obtained from imaging cells.

Fragments of data preparation and pre-formatting (S language):

```
>tmp.frame <- Generic.Summary
>names1 <- paste("Cell.line.5", tmp.names, sep=".")
> by.compound.matrix <- as.matrix(arranged.by.compound)
```

Example of the code for principal component analysis (data preparation) using S language:

```
all.data.princomp <- menuPrincomp(data =
by.compound.matrix, scores = T, cor = "Correlation",
na.action = T, print.short = T, print.importance = T,
print.loadings = T, cutoff.loadings = 0.1,
plot.screeplot = T, plot.loadings = T, plot.biplot = T,
plot.biplot.choices = c(1,2), predict.p = F)
```

Example of clustering using a divisive hierarchical clustering algorithm:

```
> div.hier.2.manhattan.cluster$call
diana(x = tmp.sum.by.comp, diss = F, metric =
"manhattan",
   stand = T, save.x = T, save.diss = T)
```

Another embodiment utilizes existing tools for biological sequence similarity searches, classification, and phylogenetic analysis . In a particular embodiment, numbers in a numerical descriptor can be substituted by one or more of nucleic acid or amino acid codes according to a one of several sets of rules. Once

converted into a corresponding nucleotide or amino acid sequence representation, the fingerprints can be analyzed and compared using software and algorithms known in the art for genetic and peptide sequence comparisons, such as GCG, a product of Genetics Computer Group, with company headquarters in Madison WI. Select embodiments comprising such approaches enable the use of a broad array of sophisticated algorithms to compare, analyze, and cluster gene and protein sequences. Many programs performing this task are known to those of ordinary skill in the art, such as for example, the PHYLIP (PHYlogeny Interference Package) a package of programs for inferring phylogenies (evolutionary trees) described in (Feldenstein, J. 1996 Methods Enzymol 266:418-427 and Feldenstein, J. 1981 J. Mol. Evol. 17(6):368-376).

Embodiments can perform such analysis based upon factors such as numerical value, statistical properties, relationships with other values, and the like. Further details of a step of manipulation are noted more particular below.

Fig. 7B illustrates a representative block flow diagram of simplified process steps for determining one or more descriptors of a state in the portions of the cells in the presence of the manipulation of step 704 of Fig. 7A in a particular embodiment according to the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. In a step 712, an image of a cell portion is obtained. In some embodiments, the cell portion is visualized with a fluorescently labeled marker that is specific for the portion or portions of interest. A cell portion can include, for example, one or more of the following: nuclei, Golgi apparatus, and other features. The cell portion may vary in select embodiments according to the invention. Then, in a step 714, a digitized representation of the image obtained in step 712 is determined. In some embodiments, steps 714 and step 712 can comprise a single step. These embodiments use a digital imaging means such as a digital camera, to obtain a digital image of the target directly. Next, in a step 716, the digital representation of the image is processed to obtain image features. Image features can include such quantities as area, perimeter, dimensions, intensity, aspect ratios, and the like. Then, in a step 718 descriptors can be determined from the image features. Descriptors can comprise scalar or vector quantities and can comprise the image features themselves, as well as

composed features, such as shape factor derived by a relationship 4π * area / perimeter, and the like. Descriptors can also comprise statistical quantities relating to feature characteristics across a population of cells, such as a standard deviation. and average, and the like.

In a preferred embodiment, cells can be placed onto a microscope, such as a Zeiss microscope, or its equivalent as known in the art. A starting point, named Site A01, is identified to the microscope. A plurality of exposure parameters can be optimized for automated image collection and analysis. The microscope can automatically move to a new well, automatically focus, collect one or more images, at one or more wavelengths, move to a next well, and repeat this process for all designated wells in a multiple well plate and for multiple plates. A file having a size and an intensity distribution measurement for each color and rank for each well can then be created for the images acquired. Based on this information, a user or a computer can revisit sites of interest to collect more data, if desired, or to verify automated analysis. In a presently preferred embodiment, image automatic focus and acquisition can be done using computer software controlling the internal Z-motor of the microscope. Images are taken using a 10x, 20x, or 40x air long working distance objectives. Sometimes multiple images are collected per well. Image exposure times can be optimized for each fluorescent marker and cell line. The same exposure time can be used for each cell line and fluorescent marker to acquire data.

Fig. 7C illustrates a representative block flow diagram of simplified process steps for obtaining images of cell portions of step 712 of Fig. 7B in a particular embodiment according to the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. The method is generally outlined by the steps below:

(1). In a step 720, a sample is provided to the imaging device. Samples can be provided in 96 well plates and the like. The sample may be loaded into a microscope, such as a Zeiss microscope or equivalent.

(2). In a step 722, a set of optical filters is selected to shine light of the appropriate wavelength to illuminate the first sample, which may be contained in a first well designated A01.

(3). In a step 724, an automatic focusing procedure is performed for the site. In a particular embodiment, the internal z-motor of the microscope which is attached to the objective nosepiece is used for automatic focusing of the microscope. In an alternative embodiments, the plate holding the samples is moved to perform automatic focusing of the microscope, or focusing can be performed by moving optical components attached to the microscope and the like.

(4). In a step 726, images are collected for the site. Images can be collected for every color at every site. Present embodiments can provide images for up to four colors. However, embodiments are contemplated that can provide more colors by using either a monochromator coupled with excitation filters which are on a filter wheel, or by digitally separating overlapping fluorophores. Those knowledgeable in the field will know that given calibration images of single fluorophores, a look-up table can be devised which will allow for the digital removal of fluorescence bleed-though of fluorescence which may occur in optical channels other than the one for which that filter has been optimized in instances of using more than one fluorophore at once. Cell growth and density information is also collected. Cell density is determined by what percentage of the area being imaged is inhabited by cells. In some embodiments, imaging can be facilitated using one or more biosensors, molecules such as non-proteins, i.e., lipids and the like, that are luminscently tagged. However, some embodiments can also use fluorescence polarization and the like. Fluorescence polarization is a homogeneous fluorescence technology where the excited state of the molecule lasts much longer than in normal fluorescence, taking seconds to minutes to reach equilibrium, obliterating the need to wash away fluorescence markers that are not specifically bound to a marker. Further, embodiments can detect differences in spectral shifts of luminescent markers. Some fluorescence markers, such as Nile Red sold by Molecular Probes of Eugene, OR, will change its emission peak wavelength depending on its environment. One can detect these changes by monitoring the level of fluorescence at both wavelengths and reading out at ratio of the two.

(5). In a step 728, a determination is made whether more fields of view need to be taken for a particular color. If this is so, then processing continues at step 726 at a new site. Otherwise, processing continues with a decisional step 730.

Images can now be taken by repeating step 726. In a preferred embodiment 4 to 9 images are collected at each site.

(5). In a step 730, a determination is made whether more optical configurations need to be taken in order to obtain images for all differently-marked cell portions the sample. If this is so, then in a step 732 a new optical configuration is determined. Images for the new optical configuration can now be taken by repeating steps 726 and 728.

(6). In a decisional step 734, after all optical configurations and images for fields of view in a sample have been obtained, a determination is made whether any further samples remain to be analyzed. If so, a new sample is brought into view and processing continues with step 720. Otherwise, image processing is complete. In a presently preferable embodiment, image data can be stored on a CD ROM using a CD ROM burner, such as CRW4416 made by Yamaha of Japan. However, other mass storage media can also be used.

Fig. 7D illustrates a representative block flow diagram of simplified process steps for processing digitized representations of step 716 of Fig. 7B in a particular embodiment according to the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. The method is generally outlined by the steps below:

(1). In a step 740, a digitized image input is preprocessed . Preprocessing might include, but is not limited to, such operations as background subtraction, thresholding, smoothing, adoptive filtering, edge enhancements, contrast enhancements, histogram equalization. A particular combination of preprocessing steps can be applied to images in successive steps or in parallel to copies of the image.

A simplified example of a smoothing and background subtraction procedure in a MatLab language is presented in computer code below:

```
function Isubtracted = cmBackgrSubtr1(I,k)
% cmBackgrSubtr1(I,k) - simple flat background (=modal*k)
subtraction
% Y = cmBackgrSubtr1(I, k) - image Y is generated by
```

```
% subtraction (with saturation) of modal pixel value of I
multiplied by k
% DEFAULT - k=1
%
if (nargin == 1)
   k=1;
end
if (size(k)~=1)
   error('cmBackgrSubtr1: parameter k should be a number.
Exiting...');
end

%modpixnum = floor(size(I(:),1)/2);
%sortedval = sort( double(I(:)) );
%modpixel = sortedval(modpixnum);
modpixel = median(double(I(:)));
bg = k*modpixel;

Isubtracted = mmsubm(  uint8(I), uint8(round(ones(
size(I))*k*modpixel )) );
```

An example of a procedure for thresholding in computer code (MatLab) is presented below:

```
function thresh = GetThreshByPerim1(I, M)
% GetThreshByPerim1(I) Finds optimal thresholding value
for image I
% N = GetThreshByPerim1(I) Finds thresholding value N for
image I
% N = GetThreshByPerim1(I, M) - tests threshold values up
to M
% DEFAULT M = maximum pixel value in I
% note that GetThreshByArea is significantly faster
% finds a threshold value that causes the maximal change
in the
```

```
% total perimeter of the objects (Russ ????)
% see Matlab_Auto_threshold1_1-23-99.doc for more details
% Note: works somewhat better on SMOOTH images (i.e.
medfilt2(I, [3 3]) two times

if (nargin == 0)
   error (strcat( mfilename,' : at least one parameter
required'));
elseif (nargin == 1)
   M = double(max(I(:)));      %test thresholds up to
maximum pixel value in I
elseif (nargin > 2)
   error (strcat (mfilename,' : too many parameters'));
end

if (size(M)>1)
   error (strcat(mfilename, ' : argument M should be a
number'));
end

Minval = double( min(I(:)));
step = 1;

%generate vertical vector perims with total perimeters of
objects at different
%threshold values
for i=Minval : step : M
     bwI = im2bw(I, i/255);
     prI = bwperim(bwI);
     pr = sum(prI(:));
     if (exist('perims', 'var') == 0) %perims is yet
undefined
          perims = pr;
     else
```

```
            perims = cat(1, perims, pr);
        end
    end

% vector prdiffs contains differences between successive
perimeters
prdiffs = diff(perims);
mindecrease = min(prdiffs);
minvalues = find(prdiffs == mindecrease);
index_of_mindecrease = minvalues(1);
thresh = index_of_mindecrease + 1;

% ==============end GetThresh1==================
```

Thresholding provides a specific intensity, such that pixels darker than the threshold are deemed black, and pixels lighter than the threshold are considered white. The thresholded image can be processed using binary image processing techniques in order to extract regions.

(2). In a step 742+744, the digitized image input is subjected to object identification. This can be accomplished by a variety of procedures, for example by thresholding or edge detection and subsequent morphological opening and closing. Edge detection can be accomplished by means of gradient-based or zero-crossing methods, such as Sobel, Canny, Laplassian, Perwitt, and other methods.

An example of object identification procedure based on Canny edge detection (in MatLab language) is presented below:

```
function Imask = cmMaskDNA1(I);
% cmMaskDNA1 - generates binary mask for cell nuclei
through edge detection
% Imask = cmMaskDNA1(I)
% PARAMETERS
%    I - intensity image (grayscale)
% OUTPUT
%    Imask - BW image with objects from I
```

```
%
% For more details see Notebook Matlab_DNA_masking1_1-22-
99.doc
% Uses SDC Morphology Toolbox V0.7

if (nargin ~= 1)
   error('Wrong number of input parameters');
end
if (nargout ~= 1)
   error('Wrong number of output parameters: one output
argument should be provided');
end


Imask = edge(I, 'canny');
Imask = mmdil(Imask, mmsecross(1));
Imask = mmero ( mmclohole(Imask,mmsecross(1)));
Imask = mmedgeoff(Imask, mmsecross(1));
% note that mmedgeoff this command removed FILLED OBJECTS
but not touching OUTLINES.
% these outlines can be removed by filtering:
Imask = medfilt2(Imask, [5 5]);

%==========end cmMaskDNA1
==================================
```

However, embodiments can also use other techniques, such as Fast Fourier Transforms (FFT) and the like as known in the art without departing from the scope of the present invention.

(3). In a step 746, a plurality of region features can be determined. For example, in a representative embodiment, image features can include such quantities as area, perimeter, dimensions, intensity, aspect ratios, and the like. Features not directly related to individual objects are also being extracted.

An example of a procedure for extraction of some of the features (MatLab language) is presented below:

```
function OData = cmGetObjectsData(I, Ilabel)
% cmGetObjectsData returns array measurements of objects
in image "I" masked by "Ilabel"
% EV 2-3-99; 2-10-99
% OData = cmGetObjectsData(I, Ilabel) returns an array of
morphological and intensity measurements
%    taken from a grayscale image "I". Objects are
identified on a mask image Ilabel, usually
%    created by bwlabel()
% OUTPUT:
% Each row in the output array OData represents
individual object
% columns  contain the following measurements:
%
%    1 - Index ("number" of an object);      8  -
Solidity;
%    2 - X coordinate of the center of mass; 9  - Extent;
%    3 - Y coordinate          -"-           ; 10 - Total
Intensity;
%    4 - Total Area (in pixels);             11 - Avg.
Intensity;
%    5 - Ratio of MajorAxis/MinorAxis;       12 - Median
Intensity;
%    6 - Eccentricity;                  13 - Intensity of
20% bright pixel
%    7 - EquivDiameter;                 14 - Intensity of
80% bright pixel
%
% For details on morphological parameters see information
on MatLab imfeature();
```

```
% Intensity parameters are either obvious or are
documented in comments in this file.

if (nargin ~= 2)
   error ('function requires exactly 2 parameters');
end
if (nargout ~= 1)
   error ('function has 1 output argument (array X by
14)');
end

% finished checking arguments

% first collect morphological parameters in a structure
array:
ImStats = imfeature(Ilabel, 'Area', 'Centroid',
'MajorAxisLength',...
   'MinorAxisLength', 'Eccentricity', 'EquivDiameter',
...
   'Solidity', 'Extent', 8 );

% now convert it into array (matrix) while collecting
intensity data for each object:

%preallocate output array:
numobjects = size(ImStats, 1);
OData = zeros(numobjects, 14);
%now convert ImStats into array and add intensity data to
it
for k=1:numobjects
     OData(k, 1) = k;
     OData(k, 2) = ImStats(k).Centroid(1);
     OData(k, 3) = ImStats(k).Centroid(2);
     OData(k, 4) = ImStats(k).Area;
```

```
        OData(k, 5) = (ImStats(k).MajorAxisLength) /
    (ImStats(k).MinorAxisLength);
        OData(k, 6) = ImStats(k).Eccentricity ;
        OData(k, 7) = ImStats(k).EquivDiameter;
        OData(k, 8) = ImStats(k).Solidity;
        OData(k, 9) = ImStats(k).Extent;

        % now collect and assign intensity parameters from
    image I

        object_pixels = find( Ilabel == k);
        object_area = size(object_pixels, 1); %same as total
    number of pixels in the object
        object_intensities = double(I(object_pixels)); %
    need to convert to double to do math
        sorted_intensities = sort(object_intensities); %
    will need to get median, 20% and 80% pixels
        total_intensity = sum(object_intensities, 1);
        avg_intensity = total_intensity / object_area;
        median_intensity = sorted_intensities( floor(
    object_area/2 ) + 1 );
        pix20 = sorted_intensities( floor(object_area*0.2)+1
    ) ; %brightest pixel among dimmest 20%
        pix80 = sorted_intensities( floor(object_area*0.8)+1
    ) ;

        OData(k, 10) = total_intensity;
        OData(k, 11) = avg_intensity;
        OData(k, 12) = median_intensity;
        OData(k, 13) = pix20; %brightest pixel among dimmest
    20%
        OData(k, 14) = pix80; %dimmest pixel among brightest
    20%
    end %for
```

```
%==================== end function
cmGetObjectsData()==============================
```

(4). In a step 748, quantitative descriptors, characterizing cell state are calculated based on the feature measurements extracted at step 746. For example, histogram distribution of intensities of cell nuclei provides information about the population cell cycle stages.

In a particular embodiment according to the present invention, data analysis techniques for describing the fluorescence patterns of cell portions in multiple cell lines in the presence and absence of compounds are provided. Automated image analysis techniques can include determining one or more regions from around nuclei, individual cells, organelles, and the like, called "objects" using a thresholding function. Objects that reside on the edge of an image can be included or excluded in various embodiments. An average population information about an object can be determined and recorded into a database, which can comprise a database text file or Excel spreadsheet, for example. However, embodiments can use any recording means without departing from the scope of the present invention. Values measured can be compared to the visual image. One or more types of numerical descriptors can be generated from the values. For example, descriptors such as a number of objects, an average, a standard deviation of objects, a histogram (number or percentage of objects per bin, average, standard deviation), and the like can be determined.

In a particular embodiment according to the present invention, data can be analyzed using morphometric values derived from any of a plurality of techniques commonly known in the art. For example, a software package called MetaMorph Imaging System, provided by Universal Imaging Corporation, a company with headquarters in West Chester, PA and NIH Image, provided by Scion Corporation, a company with headquarters in Frederick, Maryland.

Fluorescent images can be described by numerical values, such as for example, an area, a fluorescence intensity, a population count, a radial dispersion, a perimeter, a length, and the like. Further, other values can be derived from such measurements. For example, a shape factor can be derived according to a relationship

4π * area / perimeter. Other values can be used in various embodiments according to the present invention. Such values can be analyzed as average values and frequency distributions from a population of individual cells.

In a particular embodiment according to the present invention, techniques for the automatic identification of mitotic cells are provided. Image analysis techniques employing techniques such as multidimensional representations, frequency-based representations, multidimensional cluster analysis techniques and the like can be included in various embodiments without departing from the scope of the present invention. Techniques for performing such analyses are known in the art and include those embodied in MatLab software, produced by MathWorks, a company with headquarters in Natick, MA.

Scalar values providing efficacious descriptors of cell images can be identified using the techniques of the present invention to perform predictive analysis of drug behavior. In a presently preferred embodiment, a plurality of heterogeneous scalar values can be combined to provide descriptors for each manipulation. By applying predictive analysis routines to the collections of these descriptors, predictive information about any number of manipulations and cell interactions can be extracted.

Fig. 7E illustrates a representative block flow diagram of simplified process steps for analyzing image feature values to obtain descriptors of cell state of step 718 of Fig. 7B in a particular embodiment according to the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. Fig. 7E illustrates an input data of descriptors of known manipulations 319. A step 320 of reformatting and transforming data 319 to formats suitable for analysis is performed. Additionally, a "cleaning" process can eliminate outlying data points and the like in the data. Then, in a step 322, a decision is made whether to continue with step 324 or with step 326 based upon determining a particular type of analysis appropriate for the present application or particular type of prediction. If decisional step 322 determines processing should continue with step 324, then, in that step, an error estimate using a set of test descriptors is performed to estimate the quality of a prediction and processing continues with step 320. Once an optimal prediction is achieved, processing continues with step 326. In step 326, optimal transformation parameters and prediction methods are selected for use in

steps 328 and 330 which analyze data about an unknown manipulation. In a step 328, a solution is generated based upon any of techniques including training a neural network, solving a mathematical equation, applying decision tree rules and/or the like. In a step 330, an input data set of unknown descriptors 318 is reformatted and transformed based upon the optimal transformation parameters selected in step 326 using the transformation procedures in steps 320, 322 and 324. In a step 332, predictions techniques are applied to the reformatted manipulations from step 330 and the solution generated in step 328 and a plurality of properties of known manipulations 317 (e.g., therapeutic properties, and the like) in order to determine a prediction of properties of unknown manipulation 316.

Fig. 7F illustrates a representative block flow diagram of simplified process steps for a method of mapping a manipulation of cells to a physiological characteristic in a particular embodiment according to the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. The method is generally outlined by the steps below:

(1) In a step 750, a plurality of cells, e.g., dead, live, cell fractions or mixtures of cells are provided.

(2) Then, in a step 752, the plurality of cells is manipulated, where manipulation occurs using a source(s) from one or a combination selected from an electromagnetic, electrical, chemical, thermal, gravitational , nuclear , temporal , or a biological source.

(3) Next, in a step 754, a feature value is captured from the plurality of cells. The feature value can include one or any combination of characteristics such as cell count, area, perimeter, length, breadth, fiber length, fiber breadth, shape factor, elliptical form factor, inner radius, outer radius, mean radius, equivalent radius, equivalent sphere volume, equivalent prolate volume, equivalent oblate volume, equivalent sphere surface area, average intensity, total intensity, and optical density. This list is not meant to be limiting.

(4) Then, in a step 756, a degree of presence of one or more feature values is assigned for each manipulation.

(5) In a step 758, the feature values from the plurality of cells are stored in memory locations. From the memory locations the values can be used for

statistical analyses to produce predictive information about the relatedness of the descriptors of the manipulations to one another. This information is used to infer properties of the manipulations.

Fig. 7G illustrates a representative block flow diagram of a simplified process steps for a method for populating a database with manipulated biological cell information in a particular embodiment according to the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. The method is generally outlined by the steps below:

(1) In a step 760, a plurality of cells in various stages of the cell cycle, A montage image that was used as a source to generate data in Appendix A is presented in Fig. 12., such as for example, the stages of interphase, prophase, metaphase, anaphase, and telophase are provided.

(2) Then, in a step 762, each of the cells in the various stages of mitotic development is manipulated.

(3) Next, in a step 764, an image of the plurality of manipulated cells is captured using image acquisition techniques in order to provide a morphometric characteristic of each of the manipulated cells.

(4) As a preferable option, in a step 766, an image database may be populated with the image of the plurality of manipulated cells.

(5) Following step 764 or optional step 766, a morphological value is calculated from the image in a step 768.

(6) In a step 770, the database is populated with the morphological value.

Fig. 7H illustrates a representative block flow diagram of simplified process steps for a method for populating a database with manipulated biological information, e.g., image acquisition parameters, image feature summary information, and well experimental parameters in a particular embodiment according to the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. Fig. 7H illustrates a step 780 in which cells are placed into site on a plate and a manipulation is applied. Then, in a step 781 an image is taken of the cells. In step 782, the image is transferred to an image archive

database. Then, in a step 783, well experimental parameters are entered into the database 787. Well experimental parameters can include cell type, manipulation and the like. In a step 784, image acquisition parameters are transferred to database 787. Image acquisition parameters can include file name, fluorophores and the like. In a step 785, the image acquired in step 781 is analyzed. Then, in step 786, an image feature summary from the analysis step 785 is transferred to database 787.

In step 788, a lookup table for all analyses is provided to database 787. The lookup table provides information about the analyses. In a step 789, a query of database 787 for process data is performed. The results are reformatted. Then in a step 790, the database 787 is queried. Next, in a step 791, features of the manipulations stored in the database are combined and reduced. Next, in a step 793, reduced features of step 791 can be compared. In a step 792, the results of step 793 are recorded in database 787. Then, in a step 794, a report of predictions based on comparisons performed in step 793 is generated.

Fig. 7I illustrates a representative block flow diagram of simplified process steps for acquiring images of manipulated biological information, e.g., cells, cell tissues, and cell substituents in a particular embodiment according to the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. Fig. 7I illustrates a step 770 in which a user sets up an image analysis procedure. Then, in a step 772, an image is read into image analysis software. Next, in a step 774, patterns and objects are identified in the image using one or more algorithms. Next, in a step 776, sets of features are extracted from the image. Then, in a step 778, feature information, descriptor values and the like are exported to the database, such as database 787 of Fig. 7H, for recording. Next, in a decisional step 779, a determination is made whether any more images should be taken. If this is so, processing continues with step 772. Otherwise, image acquisition processing is completed.

Fig. 7J illustrates a representative block flow diagram of simplified process steps for populating, acquiring and analyzing images of manipulated biological information in a particular embodiment according to the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations,

modifications, and alternatives. Fig. 7J illustrates a step 300 of placing a plate onto an imaging stage and reading a bar code. Then, in a step 301 an autofocus procedure is performed. Next, in a step 302, a first optical filter configuration is selected and an image is collected. Then, in a decisional step 303, a determination is made whether more than one image per optical configuration can be taken. If so, then, in a step 304, a new position within the well is targeted and another image is collected. Then, in a decisional step 305, a determination is made whether any more images need to be collected. If this is so, step 304 is repeated until all images for a particular well have been collected. After one or more images are collected for the well, in a step 306, the stage is returned to a starting position within the well, and a montage is created from collected images. The results are named with a unique file name and stored.

In a decisional step 307, a determination is made whether any more optical channels in the well can be imaged. If this is so, then in a step 308 the next optical filter configuration is selected and an image is collected. Processing then continues with decisional step 303, as described above. Otherwise, if no further optical channels in the well can be imaged, then in a decisional step 309 a determination is made whether any wells remain to be imaged. If not all wells have been imaged, then in a step 310, the stage moves to the next well and processing continues with step 301, as described above. Otherwise, if all wells on the plate have been imaged, then in a decisional step 311, a determination is made whether any more plates can be processed. If this is so, then processing continues with step 300 as described above. Otherwise, in a step 312, the information is stored to a CD or other storage device as a backup.

Fig. 7K illustrates a representative block flow diagram of simplified process steps compound based upon information about effects of one or more known compounds on a cell population in a particular embodiment according to the present invention. This diagram is merely an illustration and should not limit the scope of the claims herein. One of ordinary skill in the art would recognize other variations, modifications, and alternatives. Fig. 7K illustrates a step 340 of populating a database with descriptors for known compounds. Such descriptors can be determined from imaging the cell population. However, in some embodiments, descriptors can be derived by measurements and combinations of measurements and the like. Then, in a step 342, descriptors for the unknown compound are determined from imaging a

second cell population. The second cell population has been treated with the unknown compound. Then, in a step 344, a relationship between the descriptors determined from the unknown compound with the descriptors determined from the known compounds can be determined. Finally, in a step 346, an inference can be made about the unknown compound based upon the descriptors of the known compounds from the relationship determined in step 344.

Accordingly, the present invention provides a novel database design. In a particular embodiment according to the present invention, a method for providing a database comprises measurement of a potentially large number of features of one or more sub-cellular morphometric markers. Markers can be from any of a large variety of normal and transformed cell lines from sources such as for example, human beings, fungi, or other species. The markers can be chosen to cover many areas of cell biology, such as, for example markers comprising the cytoskeleton of a cell. The cytoskeleton is one of a plurality of components that determine a cell's architecture, or "cytoarchitecture". A cytoarchitecture comprises structures that can mediate most cellular processes, such as cell growth and division, for example. Because the cytoskeleton is a dynamic structure, it provides a constant indication of the processes occurring within the cell. The cytoarchitecture of a cell can be quantified to produce a one or more scalar values corresponding to many possible cellular markers, such as cytoskeleton, organelles, signaling molecules, adhesion molecules and the like. Such quantification can be performed in the presence and absence of drugs, peptides, proteins, anti-sense oligonucleotides, antibodies, genetic alterations and the like. Scalar values obtained from such quantification can provide information about the shape and metabolic state of the cell.

In a presently preferred embodiment, scalar values can comprise morphometric, frequency, multi-dimensional parameters and the like, extracted from one or more fluorescence images taken from a number of cellular markers from a population of cells. Two or more such scalar values extracted from a plurality of cell lines and markers grown in the same condition together comprise a unique "fingerprint" or descriptor that can be incorporated into a database. Such cellular descriptors will change in the presence of drugs, peptides, proteins, antisense oligonucleotides, antibodies or genetic alterations. Such changes can be sufficiently unique to permit a correlation to be drawn between similar descriptors. Such

correlations can predict similar properties or characteristics with regard to mechanism of action, toxicity, animal model effectiveness, clinical trial effectiveness, patient responses and the like. In a presently preferred embodiment, a database can be built from a plurality of such descriptors from different cell lines, cellular markers, and compounds having known mechanisms of action (or structure, or gene response, or toxicity).

The present invention also provides database and descriptor comparisons according to other embodiments. In a particular embodiment according to the present invention, measurement of scalar values or features can provide predictive information. A database can be provided having one or more "cellular fingerprints" comprised of descriptors of cell substance interactions of drugs having known mechanisms of action with cells. Such descriptors can be compared using a plurality of techniques, such as a technique of creating "phylogenetic trees" of a statistical similarity between the descriptors from various drugs. In a present embodiment, scalar, numeric values can be converted into a nucleotide or amino acid letter. Once converted into a corresponding nucleotide representation, the descriptors can be analyzed and compared using software and algorithms known in the art for genetic and peptide sequence comparisons, such as GCG, a product of Genetics Computer Group, with company headquarters in Madison WI. In an alternative embodiment, numeric values for the fingerprints can be used by comparison techniques. A phylogenetic tree can be created that illustrates a statistical significance of the similarity between descriptors for the drugs in the database. Because the drugs used to build the initial database are of known mechanism, it can be determined whether a particular scalar value in a descriptor is statistically predictive. Finally, a compound fingerprint with no known mechanism of action can be queried against the database and be statistically compared and classified among the drugs in the database that the compound most resembles.

In a particular embodiment, relationships between measured morphometric properties and features of images and physiological conditions can be determined. Relationships can include, for example, treatment of different cell lines with chemical compounds, or comparing cells from a patient with control cells, and the like. In a presently preferable embodiment, a clustering can be performed on acquired image descriptors. Some embodiments can comprise statistical and neural

network - based approaches to perform clustering and comparisons of various descriptors. The foregoing is provided as merely an example, and is not intended to limit the scope of the present invention. Other techniques can be included for different types of data. In some embodiments, clustering and comparing can be performed on features extracted from cell images. In a presently preferable embodiment, procedures for comparisons and phylogenetic analysis of biological sequences can be applied to data obtained from imaging cells.

Select embodiments comprising such approaches enable the use of a broad array of sophisticated algorithms to compare, analyze, and cluster gene and protein sequences. Many programs performing this task are known to those of ordinary skill in the art, such as for example, the program Phylip, available at http://evolution.genetics.washington.edu/phylip.html, and other packages listed at http://evolution.genetics.washington.edu/phylip/software.html . However, select embodiments according to the present invention can comprise a technique of statistical classification, statistical clustering, distance based clustering, linear and non-linear regression analysis, self-organizing networks, and rule-based classification.

Embodiments can perform such analysis based upon factors such as numerical value, statistical properties, relationships with other values, and the like. In a particular embodiment, numbers in a numerical descriptor can be substituted by one or more of nucleic acid or amino acid codes. Resulting "pseudo-sequences" can be subjected to analysis by a sequence comparison and clustering program.

Other types of databases can also be provided according to other embodiments. The database includes details about the properties of a plurality of standard drugs. When the descriptor of a test compound is compared to the database, predictions about the properties of the test compound can be made using any known property of the other compounds in the database. For example, properties about a compound in the database could include structure, mechanism of action, clinical side effects, toxicity, specificity, gene expression, affinity, pharmacokinetics, and the like. The descriptor of a compound of unknown structure from a natural products library could be compared to the descriptors of compounds with known structure and the structure could be deduced from such a comparison. Similarly, such information could lead to better approaches to drug discovery research including target validation

and compound analogizing, as well as pre-clinical animal modeling, clinical trial design, side effects, dose escalation, patient population and the like.

According to the present invention, databases can be integrated with and complementary to existing genomic databases. Differential genomic expression strategies can be used for drug discovery using database technology. In one particular embodiment, cell data and cellular response data can be associated with a genetic expression profile assay to form a single assay. Live cells expressing fluorescence markers can be treated with a drug, imaged and analyzed for morphometry; and then analyzed for mRNA for expression. Such embodiments can provide rapid development of tools to link cellular behavior with functional genomics.

Database methods according to the present invention can be used to predict gene function and to assist in target validation. Databases that include genetic diversity, i.e., having cellular descriptors from cells of differing genetic backgrounds (tumor, tissue specific, and gene knock out cell lines), can provide the capability to compare cells of unknown genetic background to those in the database. Similarly, the descriptor of an unknown cellular portion in the presence of multiple drugs can be queried against the descriptors of the known markers in the database. For example, if an unknown gene is tagged with Green Fluorescent Protein (GFP), the database may be used to identify the cellular portions for which that unknown gene encodes.

According to the present invention, target validation and specialized cell-based assay screening can be performed using database systems and methods to serve as a universal high-throughput cell-based assay that can evaluate the molecular mechanism of drug action. As new genes are isolated and identified, a large collection of available gene-based knowledge is becoming available. From this large collection of new genes, potential protein targets can be identified using the genomic tools of sequence analysis and expression profiling. However, unless a gene mutation is tightly linked to a disease state, further validation of individual targets is a time consuming process, becoming a bottleneck in drug discovery. Furthermore, robotics and miniaturization are making "High Throughput Screening (HTS)" the industry standard, substantially reducing the time and cost of running a target-based biochemical assay. Therefore, it is now possible to routinely screen large libraries and use a resulting "hit" to validate the target. In such approaches, a specialized cell-based assay would be developed to test hits for each target. Since this often involves

the creation of cell lines expressing new markers, this stage may also become a bottleneck that cannot keep pace with HTS. In addition, these cell-based assays may not be amenable to high-throughput screening, making it difficult to test the increasing number of analogs arising from combinatorial chemistry.

In a particular embodiment according to the invention, a rapid characterization of large compound libraries for potential use as pharmaceutical products can be provided by predicting properties of compounds that relate to the compounds' potential as bioactive drugs. In many drug discovery situations, virtually millions of compounds can be passed through a HTS assay against a small number of validated targets. These assays produce hundreds to thousands of potential hits. These hits can then be subsequently screened by a pipeline of secondary and tertiary screens to further characterize their specificity, often time completely missing non-specific interactions with other proteins. Techniques according to the present invention can provide a replacement to such screening operations by providing information about cellular accessibility and mechanism of action for the hits coming from a HTS system. Furthermore, it can replace the biochemical HTS assay and allow rapid and accurate identification of attractive compounds from large libraries without an intervening biochemical assay. The cell information can be predictive of whether to continue into an animal model for each compound, and which animal model to pursue.

The principles of the present specifically contemplate a wide variety of research methodologies, or usage scenarios, implementing these principles. The following discussion of three such scenarios is by way of illustration and not limitation. Study of the principles enumerated herein will render evident to those skilled in the art certain additional methodologies or usage scenarios enabled by the teachings hereof. The present invention specifically contemplates all such modifications. The following description presents some specific embodiments and scenarios that represent a broader use of cellular phenotypic data and characterizations to deduce mechanisms of action and other features of cellular responses to various stimuli. Such procedures generally involve producing a quantitative cellular phenotype based upon two or more cellular attributes and then comparing that phenotype to phenotypes previously stored and indexed. Such

procedures make use of databases or other repositories of biological information. The invention is not limited to the specific embodiments described here.

Considering first the procedure 2000 depicted in Figure 20, a compound has been identified as having a particular cellular activity. See 2004. For example, a compound may be found to inhibit the growth of certain cancer cell *in vitro* by a specific and desired mechanism of action. This may be a particular company's "gold standard."

Next, the compound is analyzed at 2006 in terms of its effect on one or more cell lines. More specifically, the compound is linked, virtually, to a particular phenotype. Two or more values or measures of cellular attributes characterize that phenotype. These attributes are quantified in the context of specific cellular markers.

In one example, the cellular marker is an organelle such as a nucleus or Golgi apparatus. Measured attributes useful for characterizing an associated phenotype include geometric parameters (e.g., size, shape, and/or location of the organelle) and composition (e.g., concentration of particular biomolecules within the organelle).

The phenotype may be characterized by administering the compound of interest to various cell lines and in various concentrations. In each example within this matrix, the attributes of interest are measured. Ultimately, certain phenotypic features (combinations of attribute values) are associated with the compound of interest. These features provide a template for the phenotype.

Next, using the phenotype as identified at 2006, the process identifies other compounds providing similar features. The goal here is to present a list of compounds having a mechanism of action similar to that of the compound that started the process. This allows researchers to identify a mechanism of action, if not already known, for their compound and to draw conclusions based upon their compound's link to other known compounds (which may not be chemically/structurally similar to the compound of interest).

Identifying similar compounds based upon phenotype can take many paths. Most will involve some mathematical basis. For example, the phenotype defined at 2006 can be represented as a fingerprint or vector comprised of multiple scalar values of cellular attributes (as described above). The phenotype representation can then be compared against known phenotypes characterized by the same format

(e.g., they are all characterized as vectors having the same attribute set, but with different values of the attributes). The comparison may be as simple as a Euclidean distance or more sophisticated as a neural network or multivariate statistical correlation.

The known compounds and associated phenotypes may be stored as database records or other data structures that can be queried or otherwise accessed as part of the identification procedure. The compounds may also be associated with other relevant data such as clinical toxicity, cellular toxicity, hypersensitivity, mechanism of action, etc. (when available).

Compounds found to be sufficiently similar to the starting compound are returned for consideration by researchers. A data processing system may rank such compounds based on degree of similarity to the starting compound. In some cases, the system may even provide similarity scores associated with the listed compounds.

Often researchers wish to determine whether their particular compound has clinical or biochemical effects beyond those that they are already aware of. In a typical scenario, the compound of interest was selected based upon its strong binding a target or its stimulation or inhibition of cell growth in a particular cell line. The process associated with 2010 has likely identified the compound of interest as having a particular mechanism of action based on phenotypic similarity to other compounds having a similar mechanism of action. However, within the region of biochemical space, there may be subspaces (characterized by subphenotypes) that correspond to separate properties. For example, within the phenotypic space associated with one mechanism of action, there may be subspaces associated with clinical toxicity, cellular toxicity (likely overlapping the clinical toxicity space), and little or no toxicity. Obviously, a researcher would like to know whether her compound is likely to be toxic.

Thus, the process 2000 may include characterizing the compound of interest in terms of its distance from (i.e., similarity to) specific phenotypes having known characteristics. In a typical example, the known characteristic is toxicity. This feature allows the researcher to quantify her compound in terms of mechanism of action AND toxicity (or in terms of two or more other relevant properties associated

with phenotype). To allow simple ranking or characterization, compounds of interest may be scored according to a simple or weighted Boolean expression.

A second scenario of interest is depicted in Figure 21. This scenario again defines a phenotype in terms of a quantifiable vector or other measure. However, rather than using a compound of interest to generate the phenotype, some other cellular stimulus is used to generate the phenotype.

As shown, a process 2100 begins with receipt of cells of interest. See 2104. In many situations, the cells are produced by a genetic or epigenetic process that affects the expression level or activity of a particular protein. More generally, any cellular stimulus (e.g., radiation level and type, gravity level, magnetic field, acoustic perturbations, etc.) can be used to generate the cell line of interest. Importantly, this stimulus affects the phenotype and can be correlated therewith.

In the context of drug discovery, a gene encoding for a particular target can be genetically knocked out, underexpressed, overexpressed, expressed in a non-native state, etc. This may be accomplished via standard procedures involving genomic modification, translation or transcription apparatus modification (e.g., use of antisense nucleic acids), blocking target activity (using antibodies to a receptor site for example), and the like. These processes will generally affect the phenotype in some quantifiable way. Importantly, they clearly and unambiguously define a cellular phenotype associated with altering the activity of the target protein.

At 2106, the process involves measuring one or more cellular features from the cell line of interest to define/quantify the phenotype. This may be accomplished as described above with reference to 2006. Next, at 2108, the cellular phenotype generated in this manner is used to identify and rank a set of compounds associated with the phenotype. This operation may proceed in the manner of operations 2008 and/or 2010 from Figure 20.

Finally, at 2110, the process clusters the compounds returned at 2108 by a mechanism of action. The operation 2106 has tightly bound a mechanism of action to a phenotype. Various compounds characterized and stored in a system database may be tentatively assigned a mechanism of action or may have no suggested mechanism of action. By matching their virtual phenotype to the phenotype generated at 2106, one can create or strengthen an association between the compounds and mechanism of action relevant to the stimulus at 2104.

Considering now Figure 22, a third scenario is depicted. This scenario again involves using a virtual phenotype to glean information relevant to a mechanism of action or other cellular activity. In this case, assay data from a group of compounds (e.g., a primary or focused library) is used to elucidate a phenotype.

As shown, a process 2200 begins by identifying a target protein. See 2204. Then, at 2206, the process involves identifying positive and negative biochemical hits. More generally, this may involve ranking a number of compounds based upon their interaction with the target. In a specific case, the compounds are ranked based upon their binding affinities to or ability to inhibit the enzymatic activity of the target protein.

After the compounds have been characterized in some manner based upon their interaction with the target, they are used to define a cellular phenotype. See 2208. Generally, the techniques to accomplish are the same as described with reference to operation 2006 of Figure 20. In this case however, one may obtain a strong correlation between mechanism of action (involving the target) and phenotype by using multiple of the compounds identified at 2206. For example, some of the "best hits" may be administered to cell lines in various concentrations. And some of the least effective compounds may also be administered. Cellular attributes that are more strongly exhibited with increasing concentration of the best hits (and not exhibited or exhibited only weakly upon administration of the negative hits) can be used to define the virtual phenotype. In a related approach, compounds having widely varying levels interaction with the target are administered to cells. Those cellular attributes that vary linearly or at least monotonically with the degree of interaction between the target and compound represent attributes that can be used to define the virtual phenotype.

After the cellular phenotype has been defined, previously characterized compounds may be clustered with that phenotype. See 2210. As with operation 2110 of Figure 2, this may create or strengthen an association between a mechanism of action and various compounds in a database.

Finally, and optionally, procedure 2200 may provide a "higher resolution" mechanism of action for the compounds identified at 2206. See 2212. Presumably interaction with the target suggests a specific mechanism of action or at least some aspect of a mechanism of action. However, a given target may participate

in a larger cellular mechanism of action – unknown to researchers. Further, a compound may that binds with the target may participate in multiple mechanisms of action – some of which do not involve the target. By linking the target (and its positive hits) to a particular phenotype, some of these additional cellular level activities can be elucidated. The defined phenotype may have been previously identified as associated with other mechanisms of action or higher resolution mechanisms of action. Thus, the phenotype identified at 2208 can be leveraged to generate a higher resolution mechanism of action at 2212.

As suggested in the above discussion, compounds and associated phenotypes may be stored as database records. Such databases can take on many flavors. In one example, a database includes various pieces of information relevant to oncology. Such database may include numerous compounds classified by cellular phenotype, mechanism of action, toxicity, etc. More specifically, the database may include data on commercially available compounds clustered by cellular phenotypes corresponding to mechanisms of action. Further the databases of interest may extended or combined (via standard relational tables and algebra for example) to include additional data such as pharmacology data, cellular genomics data, gene expression data, protein expression data, etc. In a specific example, the database includes measurements made on a subset of the NCI60 cell lines, using DNA, Golgi apparatus, and/or microtubules as markers for defining the phenotypes. Other data includes dosage response information, variation in effect over time, etc. The compounds populating the database could include known National Cancer Institute oncology study compounds. In a specific embodiment, the compound set includes some or all of the compounds mentioned in the article "A gene expression database for the molecular pharmacology of cancer," Nature Genetics, 24, pp. 236-244 (March 2000).

Various biological analyses may be conducted to develop additional information for characterizing compound mechanisms of action, etc. For example, a cell count analysis may be used to develop dose response curves, GI 50 data, etc. The cell cycle may also be analyzed to find out how various stages in the cycle vary in response to particular stimuli. The Golgi apparatus may be analyzed to determine whether it is in a normal state, a dispersed state, a diffused state, etc. As another example, tubulin may be analyzed to determine whether it is normal, de-polymerized,

over-polymerized, bundled, etc. Obviously, combinations of such analyses may be performed. For example, properties of the Golgi apparatus or tubulin may be analyzed over one or more cell cycles.

In some embodiments, techniques according to the present invention can provide tools for the later stages of drug development such as clinical trial design and patient management. The properties of known drugs, such as clinical trial and patient response information, will be used in a similar fashion as the pre-clinical information to provide predictions about the properties of novel compounds. Because the human cell is the locus of drug action, a database containing drug-cell interactions will be able to provide predictive value for this aspect of drug development.

Although the above has generally been described in terms of specific hardware, software, and methods, it is understood that many alternatives can exist. In particular, the present invention is not limited to a particular kind of data about a cell, but can be applied to virtually any cellular data where an understanding about the workings of the cell is desired. Thus, in some embodiments, the techniques of the present invention could provide information about many different types or groups of cells, substances, and genetic processes of all kinds. Of course, one of ordinary skill in the art would recognize other variations, modifications, and alternatives. Some examples according to the present invention are provided below.

EXPERIMENTS

To prove the principle and demonstrate the objects of the present invention, experiments have been performed to determine the effects of manipulations on cell structure using imaging and analysis techniques applied to a variety of situations. These experiments were performed by growing multiple cell lines in the presence of multiple compounds, or substances. Cells were fixed and stained with fluorescent antibodies or labels to multiple cellular portions . One or more images of the cells were then obtained using a digital camera. Descriptors were built by quantifying and/or qualifying patterns of one or more feature from each image in the cell lines under study. A database was built from the descriptors. As the database grows, it should be able to predict the mechanism of action of an unknown drug by comparing its effect with the effects of known compounds or to identify data clusters within large libraries of compounds.

In a first experiment, an automated method to count the number of cells and differentiate normal, mitotic, and apoptotic cells was created. Approximately, 5,000 HeLa cells were plated per well in a 96 well plate and grown for 3.5 days. The cells were fixed with –20° MEOH for 5 minutes, washed with TBS for 15 minutes, and then incubated in 5 mg/ml Hoechst 33342 in TBS for 15 minutes. Then, 72 images were collected with a 40x objective and 75 ms exposure time.

The analysis was performed on objects that met a certain size criteria that was based on 1) measuring the size of objects in the image that were clearly not cells and 2) excluding the first peak of the area histogram (Fig. 8B values 1-4654).

Histograms of the individual object data were generated for each type of feature. Fig. 8A shows the histogram for average intensity, and Fig. 8B shows histogram data for the area of each object. Fig. 8C shows the scatter plot of the average intensity vs. the area of all of the objects. The pattern of the scatter plot showed an interesting pattern: a large cluster of cells in one region of the graph, with a scattering of object points in other regions. Because mitotic structures are identified as particularly bright objects, most likely due to the biological fact that the chromatin is condensed, the original Hoechst images could be used to identify which cells were either undergoing mitosis, or otherwise looked abnormal. Manual inspection of 917 cells resulted in the classification of each object. Fig. 8D shows a graph where each type of cellular classification is delimited. This graph clearly shows that the mitotic nuclei are brighter than the interphase nuclei. Further, the different phases of the cell cycle can be separated using these two features. Figs. 8E-8F show bar graphs of the average and standard deviations of the areas and average intensities for each cell classification type. These graphs show that interphase nuclei are statistically less bright than mitotic nuclei and that telophase nuclei are statistically smaller than other mitotic nuclei.

Each image was thresholded to an intensity level of 20. A standard area value was set at 9500 pixels. Automated information gathering about all of the objects was done and collected into an Excel spreadsheet (for more information see, section on imaging system). The following information was recorded:

IMAGE NAME
OBJECT #

AREA
STANDARD AREA COUNT
PERIMETER
FIBER LENGTH
FIBER BREADTH
SHAPE FACTOR
ELL. FORM FACTOR
INNER RADIUS
OUTER RADIUS
MEAN RADIUS
AVERAGE INTENSITY
TOTAL INTENSITY
OPTICAL DENSITY
RADIAL DISPERSION
TEXTURE DIFFERENCE MOMENT
EFA HARMONIC 2, SEMI-MAJOR AXIS
EFA HARMONIC 2, SEMI-MINOR AXIS
EFA HARMONIC 2, SEMI-MAJOR AXIS
ANGLE
EFA HARMONIC 2, ELLIPSE AREA
EFA HARMONIC 2, AXIAL RATIO
EFA HARMONIC 3, SEMI-MINOR AXIS

The following results were obtained:

- 1,250 objects were counted
- 201 of those objects has standard area counts > 2 (area > 19000 pixels)
- 195 objects had areas < 6000 pixels
- 1529 objects estimated in total
- 1328 object areas are > 6000 pixels
- The data was reduced to 917 objects that were 6000 < area < 19000
- For the 917 objects a scatter plot of area vs. average intensity and a histogram of the average intensity were generated.

- 116 objects that had average intensity intensities > 60 were manually looked at to determine their morphology.
- Of those 116 objects:

  6 were dead or indistinguishable

  4 were interphase

  30 were prophase

  32 were metaphase

  24 were anaphase

  20 were telophase (10 pairs)

- 12 prophase objects were missed because of gray scale cut off. (8 of those prophase cells had gray scale values > 57, as did 7 interphase)
- 1 telophase object was missed because it was too small (<6000)
- 1 prophase object was missed because it was too big (>1900)
- 16 mitotic objects were missed because they were parts of objects with standard count > 2.

In sum, out of 917 single objects, the analysis correctly identified 106 out of 130 mitotic objects, or (81% predictive, 91% of identified mitotics). Out of 917 single objects, the analysis incorrectly identified only 10 non-mitotics as mitotics (1% total, 8% of identified mitotics); 14 mitotics as interphase (1.4% total, 1% interphase). An automated classification system that would automatically assign values to each object using these or other measurement features can thus be developed, utilizing the principles set forth herein.

In a second experiment, the effects of Taxol on MDCK cells and the different types of morphological effects were observed. A plurality of MDCK cells grown in 96 well plates were treated with Taxol for 4.5 hours at different concentrations (10 uM-1pM). They were then fixed, labeled with Hoechst, and imaged.

This experiment used a labeling protocol comprising: MEOH fix at –20°, Wash in PBS, Block in PBS/BSA/Serum/Triton-X 100, Incubate with 5 μg/ml Hoechst 10 minutes, and wash.

Cells were inspected for different morphologies and manually counted at each different drug concentration in one well. Fig. 9 shows example images from each drug concentration and the different types of morphologies and cells are highlighted. Fig. 10 shows the distribution of each morphology within the cell population as a function of drug concentration. The higher the concentration of Taxol, the larger proportion of cells underwent apoptosis, and the fewer number of normal mitotic cells were detected.

In a third experiment, the purpose was to determine whether the automated analysis methods developed in the first experiment can detect differences in Hoechst morphology in the presence of 6 known compounds at one concentration and exposure time in one cell line. In this experiment, HeLa cells were separately treated with 6 compounds with known mechanism of action. The quantitative methods described in the first experiment were applied to the Hoechst images.

Approximately 5,000 HeLa cells per well were plated in a Costar black-walled 96 well tissue culture treated plate and left to recover in the incubator for 24 hours. After this time, 10 ug/mL of cytochalasin D (CD), Taxol, hydroxyurea, vinblastine, nocodazole, and staurosporine was added to different wells at a 1:100 addition in DMSO.

The cells were incubated in the presence of drug for 24 more hours. After 24 hours, the cells were removed and fixed as in the first experiment. Then, 9 images per well were collected of the Hoechst staining using a 10x objective.

The low magnification images taken of Hoechst were run through the automated image analysis method described in the first experiment. Plots of the average intensity and area were made of each compound. Fig. 11 shows the scatter plots of the compounds. The scatter plots of each compound are visually distinct. For example, cells treated with CD are smaller than control, and cells treated with Hydroxyurea are larger and brighter. Furthermore, the number of cells per well was very different (data not shown).

The effects of different compounds can be clearly and automatically distinguished by identifying changes in cellular morphology. This method can also be used to count adherent cells.

The next experiment was to develop clustering algorithms that assign statistically meaningful values to the representative two dimensional data shown in Fig. 10, and even more complicated clustering of all of the multidimensional data that can be extracted across one, and multiple images.

A fourth experiment was performed to obtain high magnification images of two markers in the presence of drugs. In this experiment, HeLa cells were treated with 80 generic compounds with known mechanism of action. The quantitative methods described in the first experiment were applied to the Hoechst images.

Approximately 5,000 HeLa cells per well were plated in a Costar black walled 96 well tissue culture-treated plate and left to recover in the incubator for 24 hours. After this time, 10 ug/mL of each compound from the Killer Plate from Microsource Discovery Systems (Gaylordsville, CT) was added to different wells at a 1:100 addition in DMSO. The cells were incubated in the presence of drug for 24 more hours. After 24 hours, the cells were removed and fixed as in the first experiment. In addition to being labeled with Hoechst 33342 (against chromatin), cells were also labeled with 1 unit of rhodamine-conjugated phalloidin (against actin) for 30 minutes.

The 96 well plate was imaged twice. Once, 9 images per well were collected of the Hoechst staining using a 10x objective. After this, one image per well of both the phalloidin and Hoechst staining was collected using a 40x objective.

The resulting high magnification images were analyzed qualitatively and distinct pattern differences were detected in both the Hoechst and phalloidin images. Fig. 12 shows three example images from the experiment. The top row is the Hoechst staining, and the bottom row is the phalloidin staining from the same well. The columns show the images from wells treated with just DMSO (control), cytochalasin D, and Colchicine. The morphology of each marker is different in the presence of each drug. Interestingly, there is an effect in the morphology of the chromatin in the Hoechst image of cytochalasin D, which directly targets the actin cytoskeleton (and thus there is an expected effect in the phalloidin image). Also, there is an effect on the actin cytoskeleton, compared to control, in the presence of colchicine that directly targets the microtubule network.

The low magnification images were analyzed as described in the first experiment, and different patterns were seen in both the average intensity vs. area plots, and in the number of cells per well (data not shown). Thus, changes in patterns of a marker that is "down-stream" from the direct target of a compound are detectable. Automated image analysis protocols for actin and other markers can be developed similarly, again utilizing the principles set forth herein.

A fifth experiment was performed to test quadruple labeling of 9 different cell lines grown in normal conditions. In this experiment, NCI-H460, A549, MDA-MD-231, MCF-7, SK-OV-3, OVCAR-3, A498, U-2 OS, and HeLa cells were plated. Then, the cells were fixed and stained for portions of the each cell known as DNA, tubulin, actin, and Golgi.

The following table summarizes the procedures for this experiment:

Action	Active Ingredient/Notes	Buffer	Vol/ well	Desired Time	Temp
Remove media	NOTE: gently by pipetting, not aspiration				
Fix	4% Formaldehyde	PBS	100µl	20 min	rt
Wash		TBS	100µl	5 min	rt
Wash		TBS	100µl	5 min	rt

Permeabliz e	0.1% Triton X-100	TBS	100µl	10 min	rt
Permeabliz e	0.1% Triton X-100	TBS	100µl	10 min	rt
Block	% BSA % Serum Filter sterilize before use	TBS w/azide	100µl	1hr or o/n	rt or 4°C
Primary Antibody	1:1000 dilution of DM1α	TBS + 1% BSA + 0.1% TX-100	50µl	1hr or o/n	rt or 4°C
Wash		TBS	100µl	5 min	rt
Wash		TBS	100µl	5 min	rt
Wash		TBS	100µl	5 min	rt
Fluorescent Stain	FITC lens culinaris 1:500 Rhodamine-Phalloidin 1:500 CY5 goat anti-mouse 1:100	TBS + 1% BSA + 0.1% TX-100	50µl	1 hr.	rt, dark
Wash		PBS	100µl	5 min	rt, dark
Hoechst	1:1000 dilution of 5mg/ml	TBS	100µl	15 min	rt, dark
Wash		PBS	100µl	5 min	rt, dark
Wash		PBS	100µl	5 min	rt, dark
Wash		PBS	100µl	5 min	rt, dark
Store		PBS	200µl	1 month	4°C

Cells were plated out at different densities for 48 hours. Cells were fixed and labeled by the above method. Cells were imaged using an automated imaging system that collected 9 images from each marker using a 10x objective.

Higher magnification images were collected of a few cells for demonstration purposes.

In this experiment, each cell line demonstrated different morphological patterns as determined by phase. For example, A549 cells are much more compacted than OVCAR-3 cells as determined by phase contract imaging (data not shown). The different fluorescent markers showed even bigger differences between different cell lines. Figs. 13 and 14 show 4 panels of each marker for A549 (Fig. 13) and OVCAR-3 cells (Fig. 14). The markers are Hoechst (upper left), Phalloidin (upper right), Lens culinaris (lower left), and DM1a antibody (lower right). The following table summarizes the qualitative differences between these images:

MARKER	A549	OVCAR3
Hoechst/DNA	small	large
Phalloidin/actin	fuzzy	crisp - many stress fibers
Lens culinaris/Golgi	compact	Disperse/punctate
DM1alpha Tubulin	perinuclear	evenly distributed

Higher magnification images were taken of the OVCAR3 cells. Fig. 15 shows the same markers at 20x, and Fig. 16 shows the markers at 40x. While the highest magnification images show the most detail, these images illustrate that very little morphological or feature information is lost in the 10x images.

These data exemplify the differences in morphology seen between different cell types. Thus the automated image analysis software can be customized for each marker in each cell type. Different drugs should effect these morphologies differentially.

An automated quantification method for each marker and cell line can be similarly developed.

A sixth experiment was conducted with a more sophisticated software package and to develop more flexible image recognition algorithms. In this experiment, prototype image features extraction was performed using MatLab programming language with image toolbox and SDC morphology toolboxes. Algorithms are being developed that will automatically identify objects on images and

to measure various morphological and feature parameters of these objects. Many different features for each of the cellular markers were acquired.

An example of a MatLab program called "AnalyseDNA" that takes as an input an unlimited number of images, identifies individual objects in these images based on either their intensities, or based on edge-detection algorithms, and extracts a number of morphological and intensity characteristics of these objects. A copy of this program follows:

Listing of the AnalyseDNA.m program and of some of the supporting subroutines

```
function files_analysed = AnalyseDNA(filemask, outpath,
nx, ny, filter_range, dext, modifier, sfname)
% AnalyseDNA performs measurements on files of DNA images
% V1. EV 2-11-99; 2-15-99; 2-16-99
%
% files_analysed = AnalyseDNA(filemask, outpath, nx, ny,
filter_range, dext, modifier, sfname)
%
% PARAMETERS:
%     ALL PARAMETERS ARE OPTIONAL
%
%     FILEMASK - mask for file names to be analyzed
INCLUDING PATH(for example c:\images\*.tif)
%     DEFAULT '.\*.tif' (all *.tif files in the current
directory).
%
%     OUTPATH - path to a directory where all the output
files will be placed.
%     DEFAULT - output is saved in the same directory
which contains images
%
%     NX, NY - number of individual images in montage
images along X and Y axes (DEFAULT 1)
%
```

```
%     FILTER_RANGE - 3 col-wide array (or[]). Specifies
how data is filtered when summary is calculated
%     this parameter internally is passed to GetDNAData
and then to GetSummaryData - see these
%     functions for details. For example: [2 2 Inf; 6 100
8000] will case all raws of data for which
%           values in column 2 are less than 2 and all raws
where values in column 6 are less than 100 or
%           more than 8000 to be excluded from all
calculations of a summary.
%     DEFAULT - [] (means do not filter, summarize all
data)
%
%     DEXT - string. Extension for data files being saved.
%     DEFAULT 'dat';
%
%     MODIFIER - this modifier is 'SUMMARY', summary file
is created;
%           'SUMMARY ONLY' - only summary is generated,
data for individual files are not saved
%
%     sfname - string. File name of a summary file
%     DEFAULT 'summary[date].dat'
%
% OUTPUT:
%
%     AnalyseDNA works on image files or montages. For
each image file it creates a tab-delimits file of
measured
%     parameters of all the objects in the montage with
the same base name as a montage file and extension
specified
```

```
%     by dext parameter (or .dat by default) and file
'errors[date].err' - with the list of files that matched
the
%     filemask but could not be processed.
%     If 'summary' or 'summary only' modifier is
specified, it also creates a single file
'summary[date].dat' (or
%     different extension, if specified by DEXT) which
contains summary information for all analyzed files.
%
%     ALL OUTPUT FILES are saved in a directory specified
by OUTPATH parameter
%
%     RETURNS *files_analysed* - number of files that have
been successfully processed.
%
%     Column designations in the output files are
described in GetDNAData
%
% FILE NAME CONVENTIONS
%     AnalyseDNA attempts to identify a number for each
file to identify the file in summary output.
%     It does that by looking for the first space or
underscore, followed by a number and then takes
%     as many successive numbers as it can find. If it
fails to identify a number it assigns a
%     default which is -1
%
%
% SEE ALSO GetDNAData, GetSummaryData
%
% TO DO   improve error handling in opening and writing
files (GLOBAL error_file ?)
```

```
%          include procedures for writing text headers
into the output files

if nargin > 8
   error ('Wrong number of input parameters');
end
if nargout >1
   error ('Wrong number of output parameters: only one
allowed');
end

% set defaults
need_summary = 0;
summary_only = 0;
use_default_outpath = 0;
datestring = datestr(floor(now));
if nargin == 7       % set default summary file name
   sfname = ['summary' deblank(datestring)]; % extension
will be appended later based on dext
   if  deblank(upper(modifier)) == 'SUMMARY'
      need_summary = 1;
   elseif deblank(upper(modifier)) == 'SUMMARY ONLY'
      need_summary = 1;
      summary_only = 1;
   else
      error (['Wrong parameter: unknown modifier '
modifier]);
   end
end

if nargin == 5
   % default data file extension
   set dext = 'dat';
end
```

```
    if nargin == 4
       % default filter range
       filter_range = [];
    end
    if nargin == 3
       ny = 1; % default number of images in montage along Y
    end
    if nargin == 2
       nx = 1;
    end
    if nargin == 1
       use_default_outpath = 1;
    end
    if nargin == 0
       filemask = '*.tif'
    end
    % check parameters
    if ( ~ischar(filemask) |  ~ischar(dext) | ~ischar(sfname)
    )
            error('Wrong parameter type: filename, filepath,
    dext and sfname should be strings');
    end
    if ( ( size(nx) ~= [1 1] ) | ( size(ny) ~= [1 1] ) )
       error ('Wrong parameter type: nx and ny should be
    scalars (1x1 arrays)');
    end
    if (~isempty(filter_range) & size(filter_range, 2) ~= 3)
       error ('Wrong parameter type: filter range should be
    [] or 3 - cols-wide array');
    end
    % end testing parameters

    % Generate list of files to process
```

```
datapath = getpath(filemask);
if use_default_outpath == 1
   outpath = datapath;
end
if exist(outpath, 'dir') ~= 7
   error(['Path ' outpath, 'not found. Exiting..']);
elseif exist(datapath, 'dir') ~= 7
   error(['Path ' datapath, 'not found. Exiting..']);
end

sfname = makefullname(outpath, sfname, dext);
if need_summary == 1
   if exist(sfname, 'file')
      disp(['File ', sfname, 'already exists!']);
      input ('Press ^C to abort, Enter to delete and
continue');
      delete(sfname);
   end
end


flist = FileList(getfname(filemask), datapath);
numfiles = size(flist, 1); % total number of files to
process
disp(['About to process ', num2str(numfiles), ' files']);
%DEBUG  - commented out "input" to run from Wrod
input('Press ^C to abort, Enter to continue');

% main loop where the job gets done:
error_file = makefullname(outpath, ['error' datestring
'.err']);
num_processed = 0;
num_error =0;
```

```
for i = 1:numfiles
     % first generate file name for a data output file
   current_fullname = flist(i, :); % full name with path
and extension
   current_datafile = makefullname(outpath,
makefname(getbasefname(current_fullname), dext) );

   %extract number from a filename
   fnumber = getfilenumber(current_fullname);

   % load an imagefile, record errors
   read_error = 0;
   try
      I = imread(current_fullname);
      %DEBUG
      disp (['Image file #', num2str(fnumber), '
loaded']);
   catch
      % record file-opening error in an error_file
      read_error = 1;
      num_error = num_error +1;
      msg = [current_fullname ': ' lasterr];
      add_error_msg(error_file, msg);
   end

   % extract and write data to a file in outpath
   if read_error ~=1
      if (need_summary == 0)
         %DEBUG
         disp (['Starting analysis of file #',
num2str(fnumber), '.']);
         current_data = GetDNAData(I, nx, ny, fnumber);
         %DEBUG
```

```
            disp (['Finished analysis of file #',
    num2str(fnumber), '.']);
            %load current_data.mat 'current_data';
            write_data(current_data, current_datafile);
        else      %summary needed
            %DEBUG
            [current_data, current_summary] = GetDNAData(I,
    nx, ny, fnumber, filter_range);
            %load current_data.mat 'current_data';
            %load current_summary.mat 'current_summary';
            write_summary (current_summary, sfname);
            if summary_only ~= 1
               write_data(current_data, current_datafile);
            end
        end
    end
end % of the main for loop
num_processed = numfiles - num_error;

%======================end function AnalyseDNA()
=====================================

%=================================================
================================
function result = add_error_msg(filename, msg)
% adds string MSG to an errorfile FILENAME
% returns 1 if success, 0 if failure

err_FID = fopen(filename, 'at');
if err_FID == -1
   warning(['Can not open error file ' filename]);
else
   fprintf(err_FID, '%s\n', msg);
   fclose(err_FID);
```

```
    end
    %=========================end function add_error_masg()
    ==================================

    %=================================================================
    ==================================
    function N = getfilenumber(fname)
    % returns the first number extracted from a file name
    (string) or -1 if fails to extract any number
    numbers = NumbersFromString( getfname(fname) ); % vector
    of all numbers encoded in the name

                                % (but not in the path, even if
    present)
    if isempty(numbers)
       N = (-1);    % return -1 if no numbers found in the
    name
    else
       N = numbers(1);
    end

    %====================== end function getfilenumber()
    ==================================

    %=================================================================
    ==================================
    function result = write_data(data_array, file_name)
    % writes data in a data_array in a tab-delimited ascii
    file.
    % result is 0 if success and -1 if failure
    % if file_name exists, overwrites it
    result = -1;
    try
       fid = fopen(file_name, 'wt');
```

```
        if fid ~= -1
            for k = 1:size(data_array, 1)
                fprintf(fid, '%g\t', data_array(k, :) );
                fprintf (fid, '\n');
            end
            test = fclose(fid);
            result = -1;
    catch
        result = -1;
    end

    %====================== end function write_data()
    ========================================

    %=================================================
    ================================
    function result = write_summary (s_vector, file_name)
    % appends summary vector s_vector to a file_name (ASCII
    tab-delimited file).
    % if file_name does not exist, creates it.
    % result is 0 if success and -1 if failure
    %
    result = -1;
    try
        % debug
        fid = fopen(file_name, 'at');
        result = fprintf(fid, '%g\t', s_vector);
        result = fprintf(fid, '\n');
        result = fclose(fid);
        result = 0;
    catch
        result = -1;
    end
```

```
% ==================== end function write_summary()
==================================

function Data = GetObjectsData(I, Ilabel)
% GetObjectsData returns array measurements of objects in
image "I" masked by "Ilabel"
% EV 2-3-99; 2-10-99
% OData = GetObjectsData(I, Ilabel) returns an array of
morphological and intensity measurements
%     taken from a grayscale image "I". Objects are
identified on a mask image Ilabel, usually
%     created by bwlabel()
% OUTPUT:
% Each row in the output array OData represents
individual object
% columns  contain the following measurements:
%
%     1 - Index ("number" of an object);      8  -
Solidity;
%     2 - X coordinate of the center of mass; 9  - Extent;
%     3 - Y coordinate          -"-           ; 10 - Total
Intensity;
%     4 - Total Area (in pixels);             11 - Avg.
Intensity;
%     5 - Ratio of MajorAxis/MinorAxis;       12 - Median
Intensity;
%     6 - Eccentricity;                       13 - Intensity of
20% bright pixel
%     7 - EquivDiameter;                      14 - Intensity of
80% bright pixel
%
% For details on morphological parameters see information
on MatLab imfeature();
```

```
% Intensity parameters are either obvious or are
documented in comments in this file.
% Procedures in this file are documented in notebook file
"MATLAB Measuring Nuclei (1) 1-29-98.doc"

if (nargin ~= 2)
   error ('function requires exactly 2 parameters');
end
if (nargout ~= 1)
   error ('function has 1 output argument (array X by
14)');
end

% finished checking arguments

% first collect morphological parameters in a structure
array:
ImStats = imfeature(Ilabel, 'Area', 'Centroid',
'MajorAxisLength',...
   'MinorAxisLength', 'Eccentricity', 'EquivDiameter',
...
   'Solidity', 'Extent', 8 );

% now convert it into array (matrix) while collecting
intensity data for each object:

%preallocate output array:
numobjects = size(ImStats, 1);
OData = zeros(numobjects, 14);
%now convert ImStats into array and add intensity data to
it
for k=1:numobjects
     OData(k, 1) = k;
     OData(k, 2) = ImStats(k).Centroid(1);
```

```
        OData(k, 3) = ImStats(k).Centroid(2);
        OData(k, 4) = ImStats(k).Area;
        OData(k, 5) = (ImStats(k).MajorAxisLength) /
(ImStats(k).MinorAxisLength);
        OData(k, 6) = ImStats(k).Eccentricity ;
        OData(k, 7) = ImStats(k).EquivDiameter;
        OData(k, 8) = ImStats(k).Solidity;
        OData(k, 9) = ImStats(k).Extent;

        % now collect and assign intensity parameters from
image I

        object_pixels = find( Ilabel == k);
        object_area = size(object_pixels, 1); %same as total
number of pixels in the object
        object_intensities = double(I(object_pixels)); %
need to convert to double to do math
        sorted_intensities = sort(object_intensities); %
will need to get median, 20% and 80% pixels
        total_intensity = sum(object_intensities, 1);
        avg_intensity = total_intensity / object_area;
        median_intensity = sorted_intensities( floor(
object_area/2 ) + 1 );
        pix20 = sorted_intensities( floor(object_area*0.2)+1
) ; %brightest pixel among dimmest 20%
        pix80 = sorted_intensities( floor(object_area*0.8)+1
) ;

        OData(k, 10) = total_intensity;
        OData(k, 11) = avg_intensity;
        OData(k, 12) = median_intensity;
        OData(k, 13) = pix20; %brightest pixel among dimmest
20%
```

```
        OData(k, 14) = pix80; %dimmest pixel among brightest
    20%
    end %for

    %=================== end function
    GetObjectsData()================================


    function Imask = MaskDNA1(I);
    % MaskDNA1 - generates binary mask for cell nuclei
    through edge detection
    % EV 1-22-99; 2-6-99; 2-10-99
    % Imask = MaskDNA1(I)
    % PARAMETERS
    %    I - intensity image (grayscale)
    % OUTPUT
    %    Imask - BW image with objects from I
    %
    % For more details see Notebook Matlab_DNA_masking1_1-22-
    99.doc
    % Uses SDC Morphology Toolbox V0.7

    if (nargin ~= 1)
       error('Wrong number of input parameters');
    end
    if (nargout ~= 1)
       error('Wrong number of output parameters: one output
    argument should be provided');
    end


    Imask = edge(I, 'canny');
    Imask = mmdil(Imask, mmsecross(1));
    Imask = mmero ( mmclohole(Imask,mmsecross(1)));
```

```
Imask = mmedgeoff(Imask, mmsecross(1));
% note that mmedgeoff this command removed FILLED OBJECTS
but not touching OUTLINES.
% these outlines can be removed by filtering:
Imask = medfilt2(Imask, [5 5]);

%=========end MaskDNA1 ================================
```

Given the list of image files or montages of images as an input, this program creates an individual file for each image that contains the following quantitative measurements for all objects identified in the image:

1 - Index ("number" of an object);
2 - X coordinate of the center of mass;
3 - Y coordinate -"- ;
4 - Total Area (in pixels);
5 - Ratio of MajorAxis/MinorAxis;
6 - Eccentricity;
7 - EquivDiameter;
8 - Solidity;
9 - Extent;
10 - Total Intensity;
11 - Avg. Intensity;
12 - Median Intensity;
13 - Intensity of 20% bright pixel
14 - Intensity of 80% bright pixel

A fragment of an output for a single file, containing 9 images of cells stained for DNA and acquired with a 10x objective. A montage image that was used as a source to generate data in A is presented in Fig. 17.

The same program also summarizes measurements across many files and performs statistical analysis of the summary data. It creates a summary file with the following data:

1 - Image file number;
2 - Average object Area (in pixels);
3 - STD (standard deviation) of 2;
4 - Avg. of Ratio of MajorAxis/MinorAxis;
5 - STD of 4;
6 - Avg. Eccentricity;
7 - STD of 6;
8 - Avg. EquivDiameter;
9 - STD of 8;
10 - Avg. of Solidity;
11 - STD of 10;

12 - Avg. of Extent; 13 - STD of 11

14 - Avg. of objects Total Intensity; 15 - STD of 14

16 - Avg. of objects Avg Intensity; 16 - STD of 15

18 - Avg. of objects Median intensity; 19 - STD of 18

20 - Avg. of objects intensity of 20% bright pixel; 21 - STD of 19

22 - Avg. of objects intensity of 80% bright pixel; 23 - STD of 21

An example of summary output obtained by running AnalyseDNA against 10 montage files also is shown in Appendix B.

A seventh experiment was conducted in order to use sequence analysis algorithms to analyze features of cell images. In this experiment, HeLa cells were treated for 24 hours with several different compounds, and then fixed, and stained with a fluorescent DNA dye. One image of these cells was acquired for each of the treatments and morphometric parameters and features were measured:

Resulting measurements were arranged into a string of numbers and reduced to a pseudo- nucleic acid sequence using following rules: At any given position in the sequence a number was substituted by "t" (a code for thymidine) if its value is among highest 25% of the values at the corresponding position in the data set, "g" if it is between 50% and 25%, "c" if it is between 75% and 50%, and "a" if it belongs to lowest 25% of values. Thus one descriptor or sequence was generated per treatment as illustrated in Fig. 18.

Resulting sequences were clustered using an AlignX module commercial software package Vector NTI (http://informaxinc.com), which uses a Neighbor Joining algorithm for sequence clustering.

The resulting dendrogram is presented in Fig 18. On the dendrogram the closest "leafs" correspond to the closest pseudo-sequences. Interestingly, compounds with similar mechanisms of action cluster together on the dendrogram. Another example of the generation of pseudo-sequences and clustering is shown in Fig. 19.

In some embodiments, techniques according to the present invention can provide tools for the later stages of drug development such as clinical trial design and patient management. The properties of known drugs such as clinical trial and patient response information will be used in a similar fashion as the pre-clinical

information to provide predictions about the properties of novel compounds. Because the human cell is the locus of drug action, a database containing drug-cell interactions can be able to provide predictive information for this aspect of drug development.

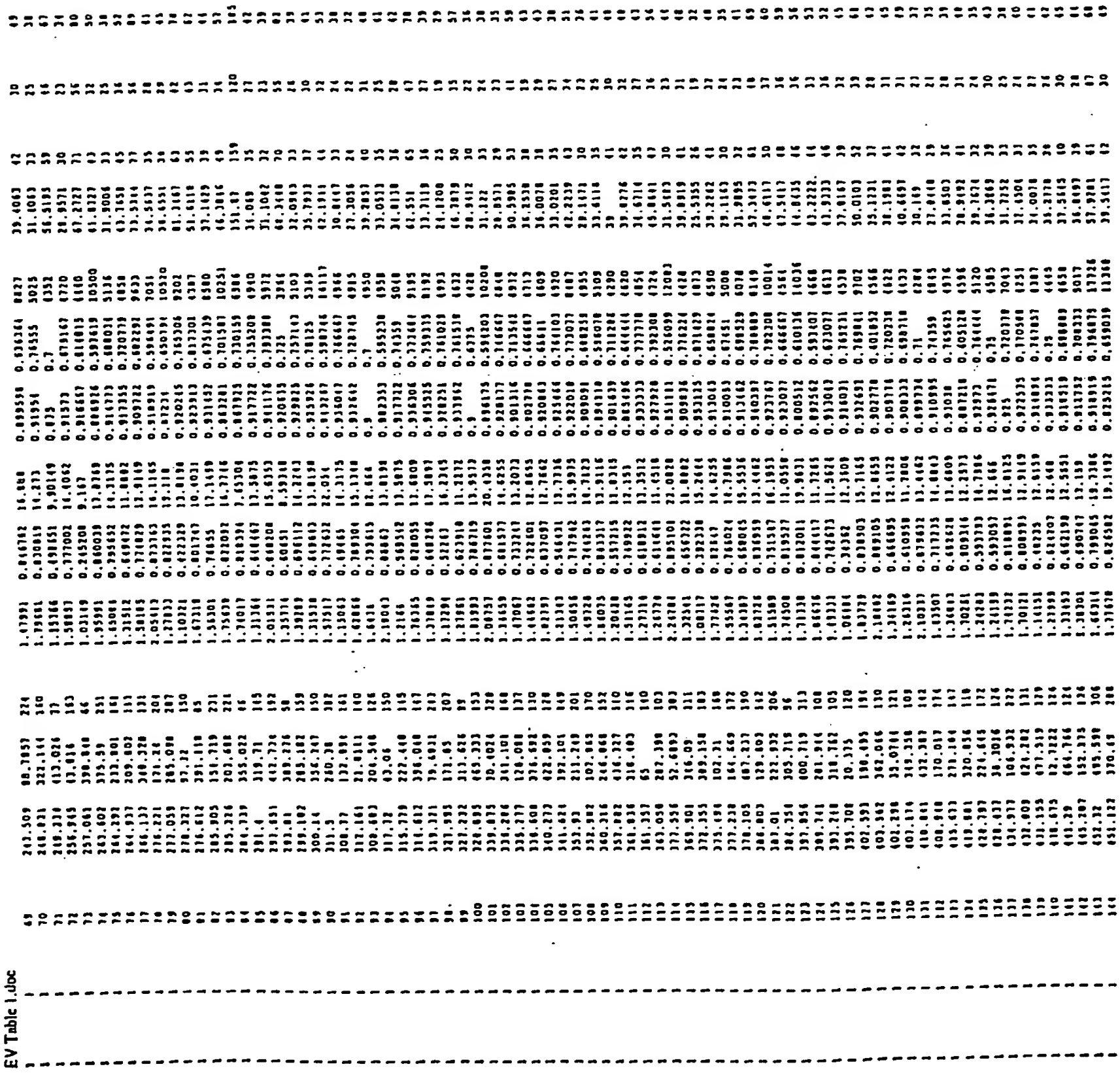
Although the above has generally described the present invention according to specific systems, the present invention has a much broader range of applicability. In particular, the present invention is not limited to a particular kind of data about a cell, but can be applied to virtually any cellular data where an understanding about the workings of the cell is desired. Thus, in some embodiments, the techniques of the present invention could provide information about many different types or groups of cells, substances, and genetic processes of all kinds. Of course, one of ordinary skill in the art would recognize other variations, modifications, and alternatives.

Appendix A

Example of the output of AnalyseDNA.m program
(measurements for a single 3 by 3 montage image)

File#	Subimage	object#	X coord	Y coord.	Area	Aspect ratio	Eccentricity	Equivalent	Solidity	Extent
1	1	1	11.2887	132.655	145	1.17293	0.322616	13.5975	0.922567	0.733756
0	-	2	16.352	416.032	125	1.60594	0.782671	12.6157	0.905393	0.70125
-	-	3	20.1073	73.8079	177	1.09815	0.413385	15.0121	0.911858	0.691106
-	-	4	21.4166	402.744	63	1.36215	0.679004	7.39828	0.911854	0.781857
-	-	5	23.0938	184	96	1.20887	0.643194	11.0558	0.888889	0.671329
-	-	6	30.3252	259.534	206	2.33106	0.903309	16.1953	0.923928	0.713278
-	-	7	33.4629	161.372	88	1.26981	0.611656	10.6451	0.927083	0.711667
-	-	8	35.0411	14.9726	116	1.25178	0.601195	12.6313	0.923936	0.711538
-	-	9	35.746	366.021	67	1.84062	0.838512	7.33578	0.81027	0.652778
-	-	10	49.1038	170.009	232	1.98691	0.851123	19.187	0.852911	0.708203
-	-	11	56.0789	126.534	221	1.95704	0.855555	16.7766	0.916686	0.688235
-	-	12	52.7755	44.9932	143	1.33627	0.683301	13.6009	0.907107	0.706731
-	-	13	53.4444	366.856	171	2.27225	0.893955	16.7555	0.871418	0.706612
-	-	14	56.4029	282.272	208	1.92782	0.858016	16.1953	0.923187	0.64375
-	-	15	51.0618	227.176	108	1.73885	0.818088	11.3265	0.933254	0.701299
-	-	16	68.1711	330.181	315	1.81138	0.832268	20.0267	0.75	0.526758
-	-	17	65.1109	402.414	220	1.70117	0.80906	16.7356	0.920502	0.667059
-	-	18	71.8669	413.13	185	1.72628	0.826503	15.3036	0.91133	0.668329
-	-	19	72.824	186.856	123	1.71588	0.811622	12.5113	0.911111	0.632231
-	-	20	77.4868	132.313	304	1.4379	0.719388	19.7388	0.932581	0.632231
-	-	21	78.7377	201.27	172	1.3357	0.682341	12.4624	0.910448	0.739384
-	-	22	81.4788	52.5012	117	1.64713	0.796614	12.2053	0.906326	0.682857
-	-	23	80.7292	281.534	213	2.13388	0.887916	21.7826	0.863281	0.731339
-	-	24	84.1765	301.926	85	1.20798	0.568991	10.4031	0.876289	0.708333
-	-	25	88.1508	126.221	113	1.43513	0.717565	12.4935	0.940789	0.781414
-	-	26	91.0529	316.926	170	1.36893	0.682889	14.7123	0.89721	0.823313
-	-	27	92.7606	211.785	288	1.87358	0.862118	18.1482	0.9	0.806316
-	-	28	96.5841	230.263	113	1.09835	0.413609	11.9568	0.896825	0.668838
-	-	29	96.8481	248.082	118	1.2774	0.622219	12.2533	0.921875	0.740234
-	-	30	101.033	93.3229	122	1.48115	0.738927	12.4631	0.915736	0.815333
-	-	31	103.47	155.507	134	1.3208	0.723411	13.0619	0.917808	0.801118
-	-	32	105.336	57.1271	110	1.90329	0.858818	12.2573	0.907692	0.801118
-	-	33	121.72	285.08	326	1.70329	0.809518	20.3734	0.900552	0.816933
-	-	34	125.332	176.645	141	1.53015	0.757813	13.3988	0.927632	0.732275
-	-	35	128.98	60.3355	152	1.75689	0.822708	13.9116	0.921212	0.737378
-	-	36	127.082	128.083	266	1.815	0.810375	18.1013	0.786607	0.801136
-	-	37	130.902	211.5	168	1.13276	0.54507	11.4303	0.932113	0.592593
-	-	38	132.439	252.545	187	1.39709	0.698312	15.4304	0.925	0.732323
-	-	39	129.813	16.6154	13	1.15921	0.50579	4.06813	0.928571	0.5625
-	-	40	134.516	209.039	181	1.12013	0.490516	11.3101	0.91069	0.617036
-	-	41	131.455	33.0909	33	1.21149	0.564506	6.48201	0.916667	0.785714
-	-	42	140.883	102.008	121	1.47767	0.736306	12.4122	0.916667	0.733333
-	-	43	144.864	59.8198	272	1.35519	0.821823	18.6087	0.916708	0.739583
-	-	44	147.093	238.035	161	1.3031	0.64397	14.2175	0.925203	0.715556
-	-	45	151.46	256.824	224	1.08868	0.317872	16.888	0.937238	0.777778
-	-	46	155.688	178.548	111	1.39135	0.695293	13.3988	0.921569	0.801136
-	-	47	160.875	202.356	68	1.31553	0.812532	7.81764	0.857143	0.592593
-	-	48	166.828	11.3677	198	1.42126	0.710535	15.0777	0.933962	0.732323
-	-	49	165.613	136.767	217	1.76118	0.820008	16.6221	0.925093	0.732323
-	-	50	176.016	356.116	222	1.46188	0.780618	16.8125	0.780036	0.611868
-	-	51	175.788	192.303	110	1.13284	0.47708	12.2573	0.900763	0.698225
-	-	52	177.181	210.678	127	1.11539	0.461601	12.7162	0.92029	0.755952
-	-	53	178.367	410.524	107	1.24127	0.582425	11.4609	0.921528	0.75
-	-	54	182.4	392.676	120	1.53768	0.755901	11.1123	0.938223	0.711806
-	-	55	188.258	242.319	136	1.81704	0.836736	15.7973	0.915888	0.702424
-	-	56	200.742	93.7118	213	1.36882	0.803828	16.4682	0.931211	0.711348
-	-	57	199.138	136.725	91	1.01662	0.235161	10.7611	0.90099	0.6
-	-	58	205.67	185.871	246	1.98093	0.860182	18.334	0.916667	0.752066
-	-	59	208.352	70.0135	230	1.69329	0.808178	17.1127	0.916902	0.714876
-	-	60	212.361	249.693	191	1.13479	0.473597	15.0376	0.928245	0.724174
-	-	61	220.056	30.6237	134	2.03562	0.923588	13.7165	0.92823	0.7124
-	-	62	216.588	234.29	183	1.36183	0.666785	15.2414	0.913239	0.699299
-	-	63	218.282	293.953	171	1.61627	0.79417	14.7355	0.918355	0.715855
-	-	64	213.331	330.721	172	1.72537	0.81181	15.7866	0.926731	0.751311
-	-	65	217.388	421.1	201	1.28372	0.634148	14.9215	0.926267	0.727273
-	-	66	222.026	151.768	121	1.3761	0.656793	12.4122	0.916667	0.733333
-	-	67	219.071	408.968	135	2.20117	0.868955	23.3262	0.76159	0.602778
-	-	68	230.344	265.186	193	1.68808	0.728893	11.3819	0.931202	0.755656

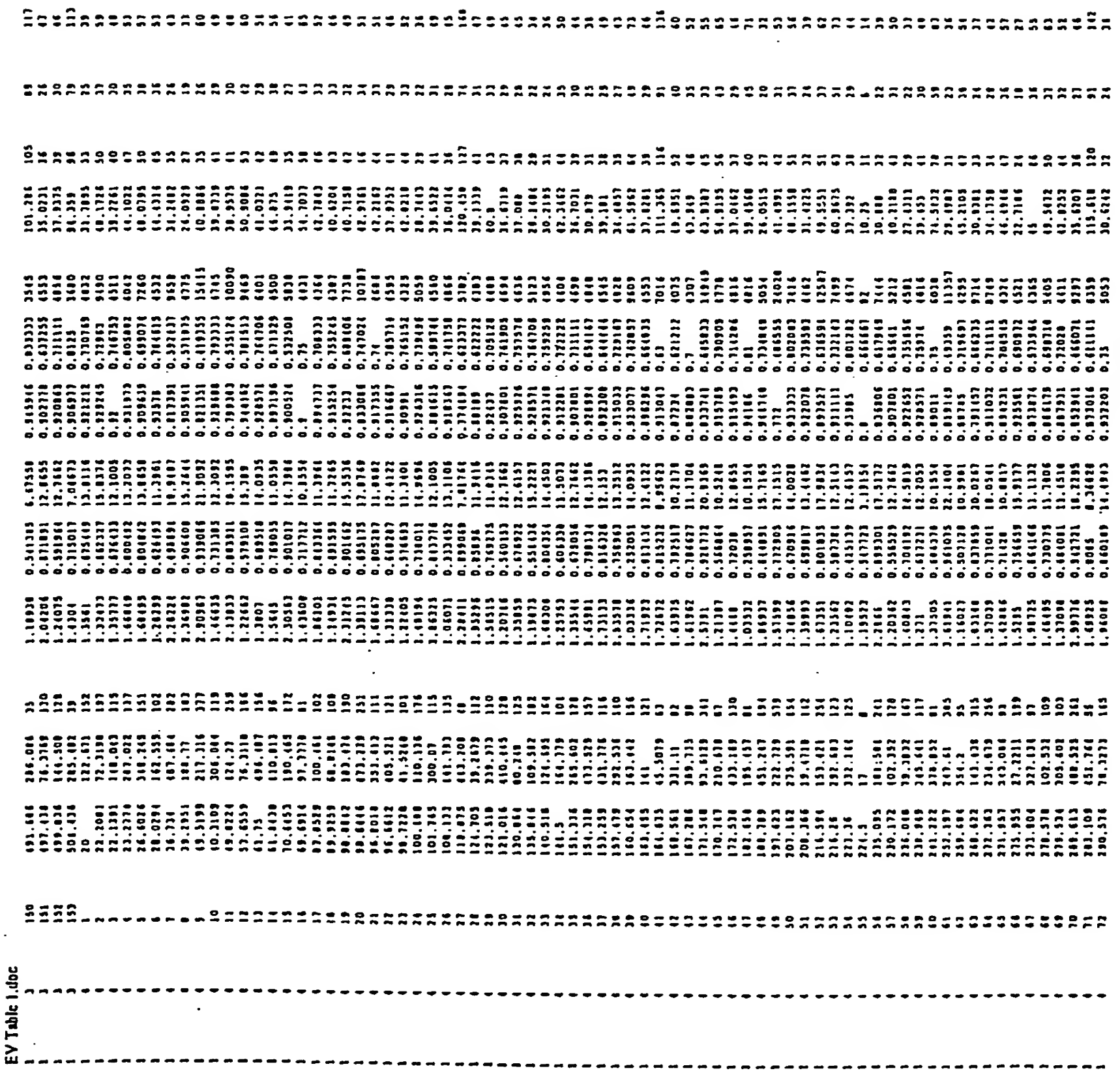
EV Table 1.doc



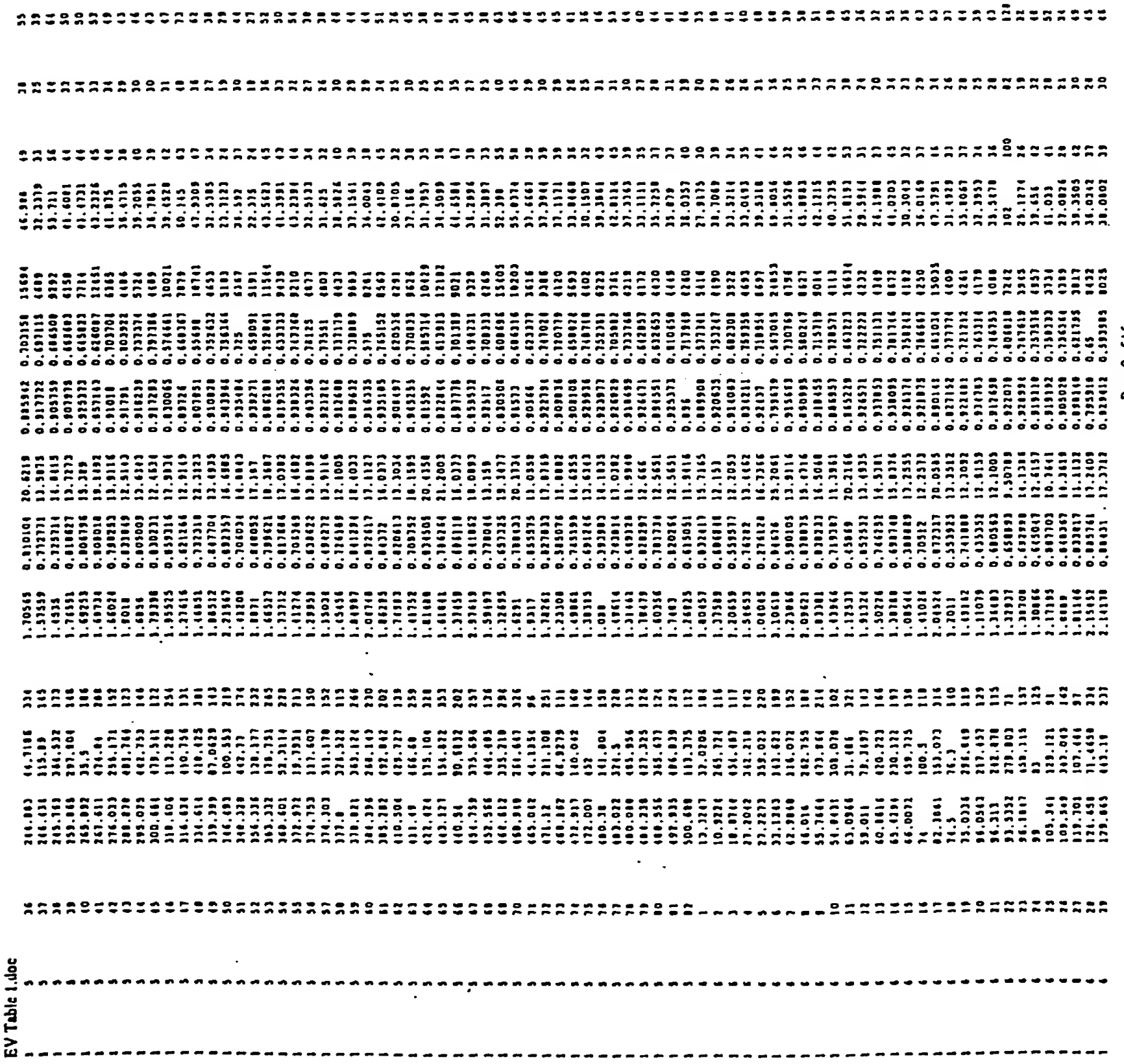
EV Table 1.doc

1	2	60	223.038	346.61	123	1.19035	0.546704	12.5143	0.926012	0.722143	4523	36.8537	38	29	44
1	2	61	231.901	273.099	171	1.81159	0.845292	16.7555	0.926324	0.7125	5607	33.2513	35	26	40
1	2	62	230.337	390.573	176	1.66435	0.793878	15.0545	0.927083	0.698039	4706	26.8836	28	21	34
1	2	63	233.936	63.4295	298	1.96264	0.850461	19.4709	0.869003	0.620833	9681	32.4066	34	26	39
1	2	64	243.615	495.715	400	1.27832	0.621307	24.7215	0.933852	0.489655	23344	48.8333	52	40	58
1	2	65	239.465	27.4994	159	1.59375	0.770656	14.2203	0.908571	0.660067	6050	30.0503	39	29	47
1	2	66	238.765	118.85	107	1.3017	0.630175	11.672	0.90678	0.748252	5042	47.1215	49	39	56
1	2	67	248.411	107.352	219	1.76031	0.822971	16.6585	0.8327	0.608333	10494	47.9178	50	38	59
1	2	68	249.387	249.387	161	1.1782	0.528789	14.4062	0.936782	0.77619	6902	29.6601	31	22	35
1	2	69	251.401	459.551	147	1.36285	0.679415	13.6809	0.926528	0.745625	5259	35.1755	37	28	44
1	2	70	252.144	48.2734	139	1.28227	0.625943	13.3034	0.914474	0.722222	6892	49.5027	51	39	60
1	2	71	265.238	469.394	160	1.46715	0.731733	14.273	0.926455	0.784314	5318	33.2375	35	28	38
1	2	72	210.909	422.169	516	3.69316	0.962643	25.6338	0.786505	0.565789	26246	46.9881	48	36	60
1	2	73	273.05	102.567	201	2.48028	0.915121	15.9975	0.922018	0.666231	5194	25.8408	27	19	32
1	2	74	275.753	332.912	170	1.84216	0.839035	14.7123	0.918919	0.708333	6885	28.7588	30	21	37
1	2	75	282.79	24.8535	152	1.65739	0.727454	14.1386	0.94012	0.769608	8307	56.7375	59	46	66
1	2	76	306.73	201.073	178	1.19701	0.549523	15.0545	0.92228	0.747059	9829	55.2191	56	41	68
1	2	77	305.481	481.562	185	2.0865	0.877615	15.3676	0.91133	0.629252	4596	20.8432	26	19	30
1	2	78	307.247	40.9837	158	1.26996	0.599963	14.1835	0.936911	0.759615	5031	31.8418	34	26	37
1	2	79	308.276	329.299	127	1.83279	0.838036	12.7162	0.900709	0.661458	4575	36.0236	37	27	44
1	2	80	310.67	144.858	176	1.86098	0.84682	14.9696	0.931217	0.651852	6921	27.9502	29	21	34
1	2	81	315.681	247.286	229	1.52513	0.753038	17.0755	0.923307	0.726984	10311	45.0262	47	34	55
1	2	82	316.476	302.312	170	1.42819	0.713958	14.7123	0.923913	0.714286	6865	28.6176	29	23	35
1	2	83	320.661	216.509	229	1.37765	0.687826	17.0382	0.930272	0.76	10187	44.4798	46	37	53
1	2	84	320.521	423.867	226	2.36105	0.956716	16.9633	0.896825	0.436615	6603	20.3613	21	15	25
1	2	85	319.458	376.38	216	2.73428	0.930722	16.5837	0.892562	0.84	4741	21.9491	23	17	27
1	2	86	323.423	62.6264	182	1.43137	0.715481	15.2227	0.923858	0.722222	9774	53.7033	51	43	63
1	2	87	340.016	384.683	246	2.34015	0.904059	17.6978	0.81457	0.630769	9905	40.2642	43	32	49
1	2	88	344.4	270.337	285	1.43643	0.717877	19.0492	0.931373	0.714286	10324	36.2246	37	29	44
1	2	89	340.84	65.4787	98	1.19007	0.542146	10.94	0.903846	0.712121	4484	47.4894	51	37	57
1	2	90	341.06	478.91	174	2.01064	0.867537	13.0619	0.937063	0.783626	4299	32.0021	33	25	39
1	2	91	345.369	334.047	149	1.10502	0.425497	13.7736	0.925466	0.760204	4161	31.953	33	24	39
1	2	92	352.892	371.092	249	1.88813	0.840967	17.8055	0.932584	0.691667	6568	26.3775	27	21	33
1	2	93	354.705	391.672	122	1.46766	0.731944	12.4634	0.917293	0.67033	9313	76.3361	82	63	89
1	2	94	353.903	77.2419	124	1.57031	0.773669	12.5651	0.892086	0.688889	4741	30.2239	40	30	46
1	2	95	355.409	136.509	110	1.44194	0.720445	11.8343	0.916667	0.705714	4501	40.9162	42	32	49
1	2	96	361.432	208.768	196	1.20688	0.559863	15.7185	0.932692	0.760784	7905	41.1598	43	32	50
1	2	97	363.681	417.549	113	2.23659	0.930843	11.9948	0.849624	0.523188	2950	28.1062	26	16	36
1	2	98	370.626	159.489	131	1.56113	0.769945	12.9149	0.919275	0.81875	4759	36.3282	37	29	43
1	2	99	373.978	497.422	225	1.84477	0.840332	16.9757	0.918367	0.669643	9414	41.84	43	33	52
1	2	100	375.192	63.568	125	1.25276	0.673444	12.6157	0.932836	0.739685	4708	37.664	39	30	46
1	2	101	372	428.5	6	1.5	0.745356	2.76395	1	1	54	9	9	9	11
1	2	102	375.31	461.469	113	2.32052	0.902381	11.9948	0.862595	0.827778	2648	23.4336	24	17	30
1	2	103	372	461	1	1	0	1.12038	1	1	9	9	9	9	9
1	2	104	380.145	395.421	145	1.61066	0.703919	13.5875	0.917722	0.697115	4480	30.8966	32	24	38
1	2	105	384.421	101.916	202	1.72134	0.813944	16.0373	0.874459	0.590643	8500	42.0792	42	29	54
1	2	106	385.745	314.015	195	1.58739	0.776623	15.757	0.934171	0.684211	9749	49.9949	52	38	63
1	2	107	391.915	291.683	82	1.56076	0.767781	10.2179	0.87234	0.630769	7836	95.561	100	77	111
1	2	108	407.296	388.112	119	1.93382	0.855916	15.0967	0.92268	0.639286	4942	27.6089	29	20	36
1	2	109	405.333	167.802	162	1.96376	0.860631	14.3618	0.9	0.609023	4778	29.4528	31	22	36
1	2	110	405.662	56.9615	130	1.23602	0.587746	12.8655	0.915493	0.714286	4523	34.7923	36	27	42
1	2	111	409	479.357	128	1.86558	0.844201	12.9159	0.895833	0.614286	4429	36.3333	37	26	42
1	2	112	410.228	100.444	180	1.26967	0.631482	15.1288	0.891089	0.705882	7824	43.6667	47	33	53
1	2	113	426.611	225.569	270	1.81845	0.835218	18.5612	0.891089	0.613636	9758	36.1407	38	28	44
1	2	114	423.62	401.364	129	1.41093	0.705455	12.8159	0.902098	0.716667	4596	35.6279	37	27	43
1	2	115	425.799	183.281	139	1.85664	0.842596	13.3034	0.908457	0.820536	6569	32.8705	35	25	40
1	2	116	429.85	439.218	206	1.51678	0.752647	16.1953	0.936364	0.774436	7146	36.6893	35	26	43
1	2	117	425.908	491.957	131	1.44743	0.723062	12.9649	0.916084	0.746316	6530	49.8673	53	37	60
1	2	118	436.543	50.8766	130	1.25228	0.601938	13.2555	0.926174	0.701082	4540	32.8986	35	26	39
1	2	119	439.22	366.185	218	2.32816	0.90288	16.6603	0.875502	0.605556	9123	41.8486	43	33	52
1	2	120	433.788	385.392	51	1.25129	0.601101	8.05824	0.944444	0.796875	3773	73.9804	72	60	89
1	2	121	436.309	477.414	152	1.08643	0.390878	13.9116	0.81018	0.72381	4446	29.2369	30	23	35
1	2	122	446.156	154.905	147	1.93836	0.856328	13.6809	0.93038	0.773684	4765	32.551	34	25	39
1	2	123	448.056	77.784	162	1.62655	0.788884	14.3619	0.9	0.710526	4549	28.0802	30	22	34
1	2	124	446.81	389.724	42	1.13349	0.479824	7.31273	0.893617	0.75	3402	82.9048	87	67	104
1	2	125	452.475	39.766	141	1.56502	0.769232	13.3988	0.933775	0.734375	8316	58.9787	61	47	71
1	2	126	457.992	423.558	129	1.35097	0.672379	12.8159	0.921429	0.732955	4753	36.845	38	27	46
1	2	127	457.668	491.056	124	1.31051	0.693618	12.5651	0.910519	0.751515	4109	33.1371	34	26	41
1	2	128	462.252	30.4967	151	1.62568	0.788427	13.8638	0.937888	0.725962	7643	50.6159	53	40	60
1	2	129	466.371	220.947	232	1.63718	0.791781	17.187	0.926303	0.692353	9151	39.444	41	31	48
1	2	130	470.772	37.88	250	3.90476	0.958431	17.8412	0.859107	0.636783	9444	37.776	38	29	47
1	2	131	472.497	344.745	145	1.74711	0.815994	13.5875	0.889571	0.69338	4916	33.9034	35	27	41
1	2	132	476.204	440.578	109	1.32316	0.636383	11.7806	0.923129	0.707192	6024	36.0257	39	29	43
1	2	133	475.31	436.126	113	1.32543	0.636329	11.9948	0.918699	0.733766	3961	35.0533	37	28	42
1	2	134	476.222	407.025	81	1.22039	0.573208	10.1554	0.9	0.736364	2613	29.7901	32	24	36
1	2	135	491.41	192.043	205	2.29749	0.900306	16.1553	0.931818	0.776515	5209	25.4098	26	19	32

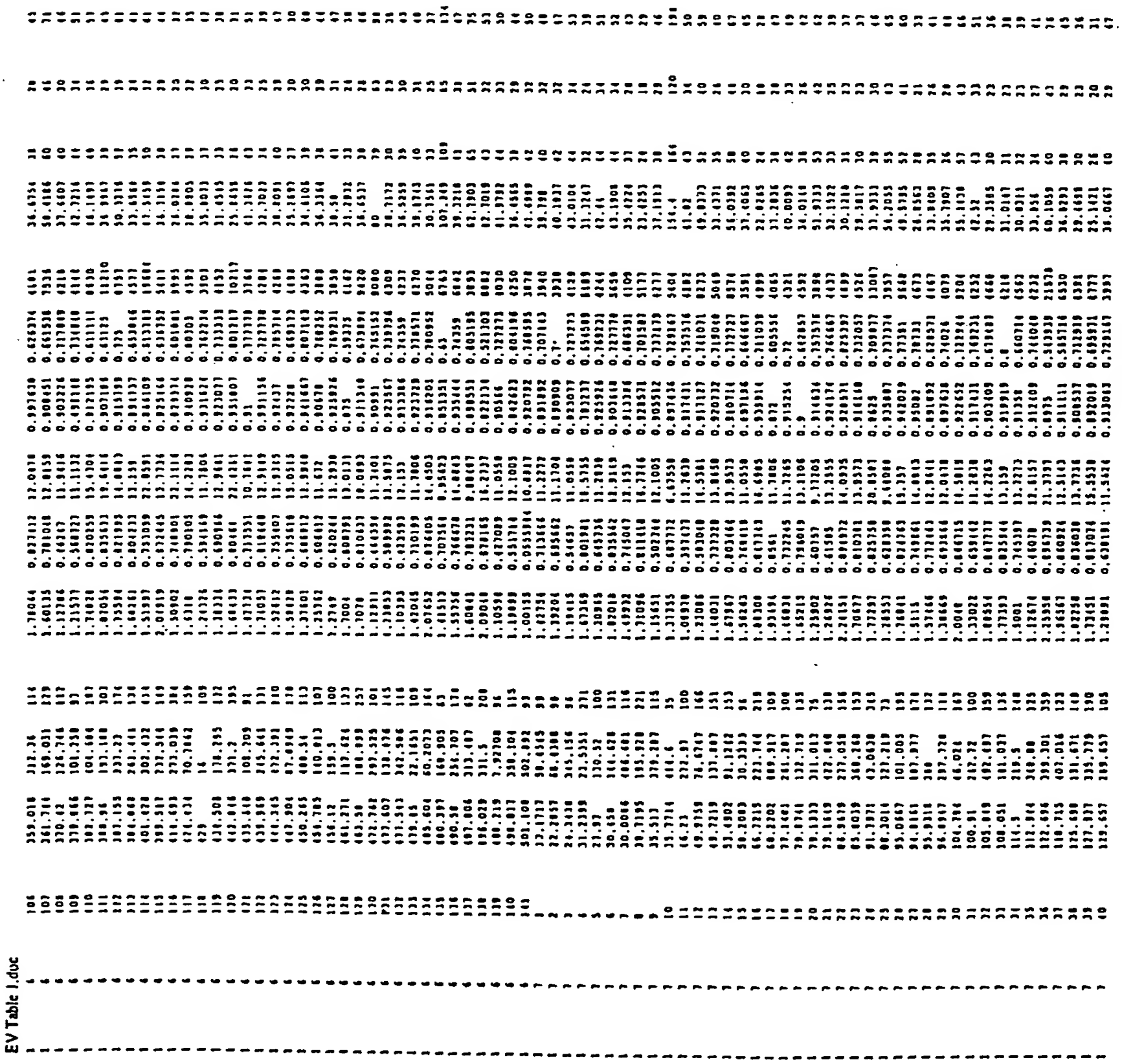
1	3	74	270.921	433.627	177	1.27986	0.624112	15.0121	0.917098	0.743697	9749	55.0791	55	42	69	
1	3	75	272.484	87.2443	221	1.7298	0.815964	16.7746	0.884	0.613889	9572	43.2122	45	33	53	
1	3	76	273.622	345.915	208	2.53252	0.91874	16.2737	0.808297	0.619048	9810	41.1435	45	36	57	
1	3	77	287.91	395.164	488	2.04761	0.872635	24.9267	0.761661	0.530435	15133	31.0102	32	24	38	
1	3	78	282.431	190.639	144	1.86664	0.644395	13.5406	0.9	0.666667	7827	51.3562	56	42	68	
1	3	79	288.361	256.084	249	2.08567	0.928094	17.8055	0.778125	0.523109	10486	42.1124	43	33	51	
1	3	80	280.085	492.603	189	1.23762	0.589177	15.5128	0.921551	0.75	4743	25.0952	27	20	30	
1	3	81	283.268	23.6071	112	1.69137	0.741848	11.9416	0.888889	0.622222	4842	43.2321	47	36	59	
1	3	82	286.84	279.133	293	1.2274	0.575125	19.3167	0.951299	0.820728	13163	65.9488	67	36	56	
1	3	83	286.967	66.0877	116	1.53691	0.759372	12.0478	0.890625	0.690909	4589	40.2566	40	31	49	
1	3	84	294.161	204.621	126	1.51011	0.748323	12.5651	0.939394	0.826667	6013	48.1919	51	39	57	
1	3	85	294.434	349.69	129	1.33034	0.659521	12.8159	0.902098	0.716667	6258	48.5116	51	39	58	
1	3	86	299	473.056	234	1.32743	0.657635	17.2509	0.83871	0.724158	8552	36.567	38	28	44	
1	3	87	299.08	329.592	113	1.40776	0.783031	11.9948	0.895825	0.620879	5052	44.708	47	36	54	
1	3	88	305.782	148.492	238	1.37652	0.687158	17.4078	0.923688	0.695906	10710	45	47	35	54	
1	3	89	308.422	376.07	128	1.92602	0.85465	12.3662	0.927536	0.790123	6686	36.6084	38	28	43	
1	3	90	310.653	61.0155	323	1.64555	0.794166	20.2795	0.828205	0.621154	9695	30.0155	29	21	43	
1	3	91	317.211	458.409	473	1.69265	0.762166	26.5406	0.852252	0.615583	16175	34.1946	35	25	42	
1	3	92	310.3	258.76	100	1.66277	0.798963	11.2938	0.900901	0.714286	4835	48.75	51	38	56	
1	3	93	316.394	172.431	109	1.70657	0.810332	11.7306	0.939655	0.801101	6791	61.9561	66	50	56	
1	3	94	319.03	158.262	99	1.66078	0.719907	11.2232	0.908333	0.692308	6732	61.198	65	39	56	
1	3	95	320.597	216.666	134	1.60176	0.700769	13.0619	0.930556	0.812121	6915	36.6191	37	28	46	
1	3	96	330.547	287.906	339	1.67331	0.723071	19.3334	0.926667	0.763736	5010	38.0692	36	28	46	
1	3	97	337.579	327.023	271	2.09771	0.870442	18.5753	0.888555	0.6335	10198	33.631	45	29	46	
1	3	98	333.171	395.74	192	1.71739	0.81299	15.6153	0.868378	0.673451	10266	51.2512	56	41	66	
1	3	99	331.564	104.017	172	1.60909	0.783419	14.1966	0.926131	0.736306	8962	53.8023	60	48	69	
1	3	100	338.308	340.832	133	1.55502	0.741454	13.0131	0.923611	0.791667	5261	38.606	40	31	47	
1	3	101	342.451	380.048	321	1.60328	0.701538	22.5631	0.918519	0.661319	9215	42.9565	45	33	51	
1	3	102	351.164	335.709	179	1.78721	0.838916	15.0167	0.889683	0.619127	9618	28.3661	29	22	35	
1	3	103	348.83	261.607	141	1.21394	0.568368	14.3135	0.941058	0.835661	5106	31.7113	32	24	38	
1	3	104	355.573	192.681	206	1.75643	0.8217	24.1953	0.919643	0.66175	7196	33.8613	38	29	47	
1	3	105	362.143	171.473	357	1.5389	0.723864	21.3201	0.788636	0.696538	19362	56.2913	57	40	67	
1	3	106	365.191	39.6394	177	1.28105	0.629539	15.0121	0.931519	0.761681	6816	27.209	28	21	33	
1	3	107	364.791	39.4298	116	1.06965	0.356967	12.0478	0.932	0.730769	4533	39.1182	40	29	50	
1	3	108	364.543	150.4	120	1.4871	0.722818	12.3608	0.941682	0.8	4669	39.013	40	32	46	
1	3	109	368.913	285.978	66	1.26168	0.679896	7.65306	0.886615	0.636889	6786	167.522	151	111	181	
1	3	110	380.868	339.11	127	1.60286	0.821706	12.7162	0.913649	0.751479	6793	33.6182	40	30	46	
1	3	111	381.023	50.5496	131	1.56783	0.76328	12.9149	0.916084	0.746318	4369	33.3511	36	27	40	
1	3	112	384.055	301.782	165	1.69253	0.74236	16.4943	0.953757	0.808824	9802	59.4061	63	48	70	
1	3	113	386.369	106.795	155	2.07424	0.816114	15.757	0.933014	0.770751	5143	26.3744	28	19	33	
1	3	114	398.813	384.067	134	1.35801	0.676578	13.0619	0.937063	0.812121	6920	36.7164	39	30	43	
1	3	115	402.106	171.906	202	1.96693	0.861118	16.0333	0.918182	0.655844	5475	27.106	28	22	33	
1	3	116	402.62	214.787	239	1.7078	0.810437	17.4462	0.895131	0.663889	13837	57.6954	58	43	74	
1	3	117	407.701	469.368	117	1.49392	0.742919	12.3053	0.936	0.75874	4361	37.2735	39	28	46	
1	3	118	412.591	72.0269	186	2.25951	0.896771	15.389	0.920192	0.673913	4700	25.2688	27	19	31	
1	3	119	420.651	404.54	137	1.87173	0.845318	13.2073	0.933333	0.652381	6163	44.9454	48	36	54	
1	3	120	421.21	439.79	100	1.74123	0.818664	11.2938	0.896956	0.666667	4125	41.25	42	32	51	
1	3	121	421.593	210.642	67	1.77433	0.826053	9.23618	0.917808	0.697917	4008	59.8209	60	47	73	
1	3	122	428.803	126.659	132	1.33134	0.696287	12.5641	0.921077	0.781065	6590	36.7727	35	27	43	
1	3	123	428.429	668.596	112	1.35329	0.673829	11.9416	0.92562	0.8	4511	40.2768	42	33	48	
1	3	124	440.679	30.7668	193	1.83016	0.837524	15.6759	0.919048	0.70696	4646	24.0725	25	19	29	
1	3	125	441.736	399.443	268	1.39649	0.696963	17.6979	0.933261	0.739298	10094	41.0325	42	32	50	
1	3	126	439.733	202.893	75	1.92134	0.85398	9.77205	0.872093	0.568182	4292	57.2267	61	46	71	
1	3	127	435.908	324.937	44	1.33865	0.666673	7.48582	0.956522	0.765714	6863	155.977	161	120	193	
1	3	128	444.902	155.214	173	1.85723	0.503268	20.8435	0.920213	0.720833	4938	28.5436	30	22	35	
1	3	129	445.937	448.141	85	1.25469	0.605633	10.9961	0.904762	0.730769	4531	67.6947	68	36	59	
1	3	130	444.353	188.412	36	1.24166	0.592169	6.57352	0.85	0.708333	5614	165.118	176	108	219	
1	3	131	447.062	79.8769	130	1.18577	0.443558	12.4655	0.928571	0.77381	4603	35.4077	38	28	42	
1	3	132	449.534	47.0451	133	1.06714	0.349099	13.0131	0.904762	0.678571	4683	35.2105	36	27	42	
1	3	133	450.026	354.38	114	1.28391	0.627186	12.0478	0.942149	0.76026	4701	41.2368	42	32	51	
1	3	134	460.248	103.075	326	3.00305	0.942929	16.8933	0.764322	0.427221	8764	38.7788	31	24	42	
1	3	135	451.634	476.634	71	1.27497	0.623033	9.50185	0.910256	0.788889	4160	58.5915	60	48	70	
1	3	136	463.959	250.088	168	1.36367	0.619888	13.7273	0.936709	0.758974	5502	37.1357	38	29	41	
1	3	137	466.043	56.1739	33	1.31472	0.649196	5.61152	0.851852	0.657143	2436	114.522	118	83	149	
1	3	138	465.3	63.85	20	1.20283	0.555713	5.01627	0.833333	0.666667	2430	121.5	132	87	154	
1	3	139	474.831	370.517	87	1.37483	0.644253	10.5218	0.878788	0.669231	5493	63.1379	66	51	77	
1	3	140	478.665	164.937	158	1.39125	0.655242	14.1835	0.929412	0.77451	8738	55.3038	56	42	70	
1	3	141	479.295	102.409	44	1.52203	0.751875	7.68482	0.916667	0.698413	6727	152.886	156	111	180	
1	3	142	482.725	413.554	204	1.77691	0.826609	16.1165	0.898678	0.6	9379	45.9755	48	35	59	
1	3	143	485.122	58.085	296	1.64407	0.721431	19.3477	0.863822	0.786096	17665	60.015	62	47	72	
1	3	144	482.755	453.941	102	1.91739	0.858085	11.3961	0.902655	0.700333	5666	55.549	56	37	71	
1	3	145	480.108	483.45	120	1.77476	0.826146	12.3608	0.909091	0.681818	4258	35.4833	36	27	44	
1	3	146	493.859	116.449	158	1.2289	0.581233	14.1835	0.918605	0.705357	4655	29.462	30	22	36	
1	3	147	496.56	345.393	298	1.53528	0.758603	19.6188	0.792553	0.627368	9905	33.2383	35	25	41	
1	3	148	489.322	61.1726	87	1.17066	0.520168	10.5218	0.887755	0.725	3815	43.8506	46	35	52	
1	3	149	498.5	469.406	128	1.50512	0.747379	12.7662	0.936307	0.8	4444	34.7188	35	25	43	

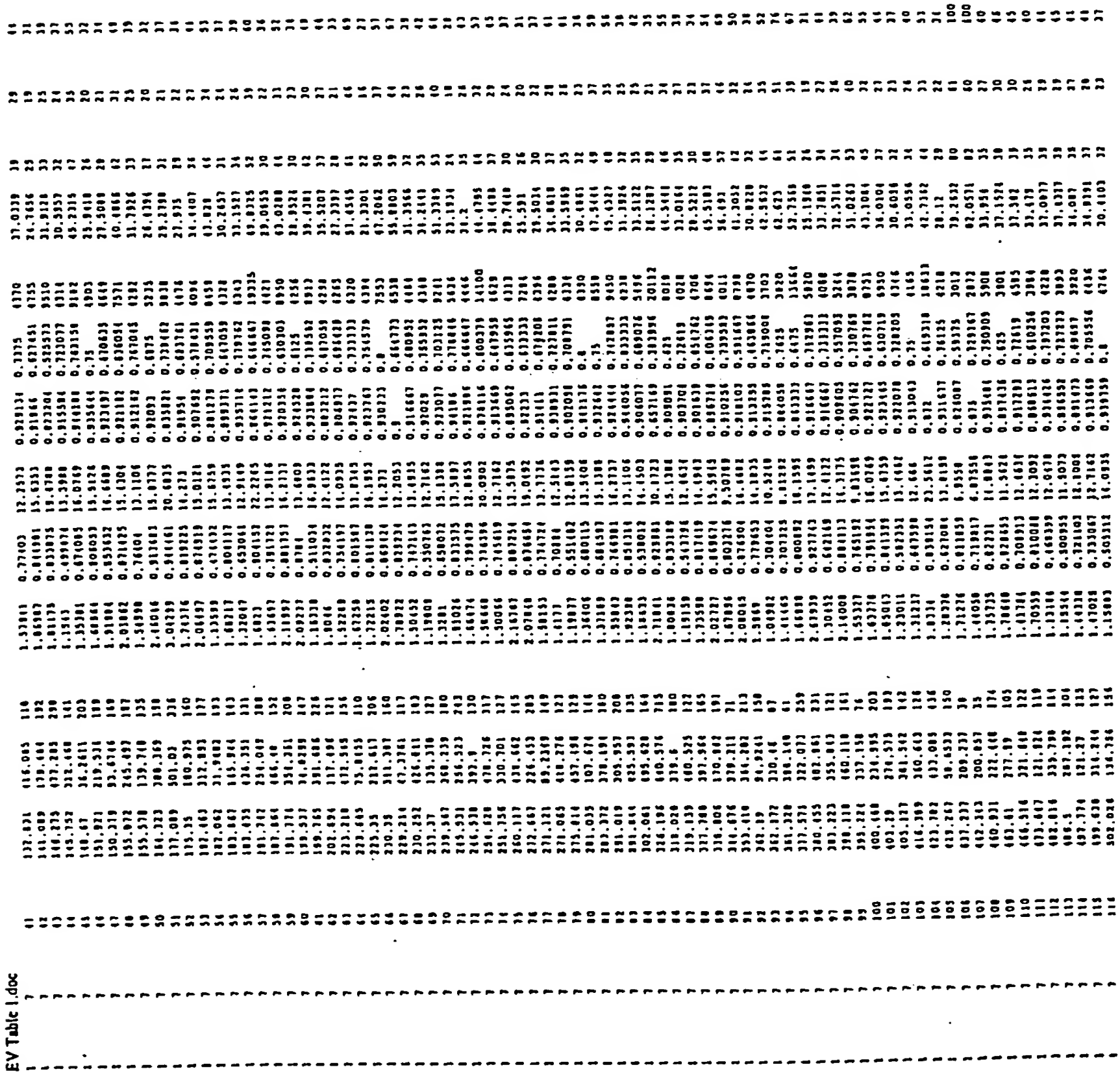


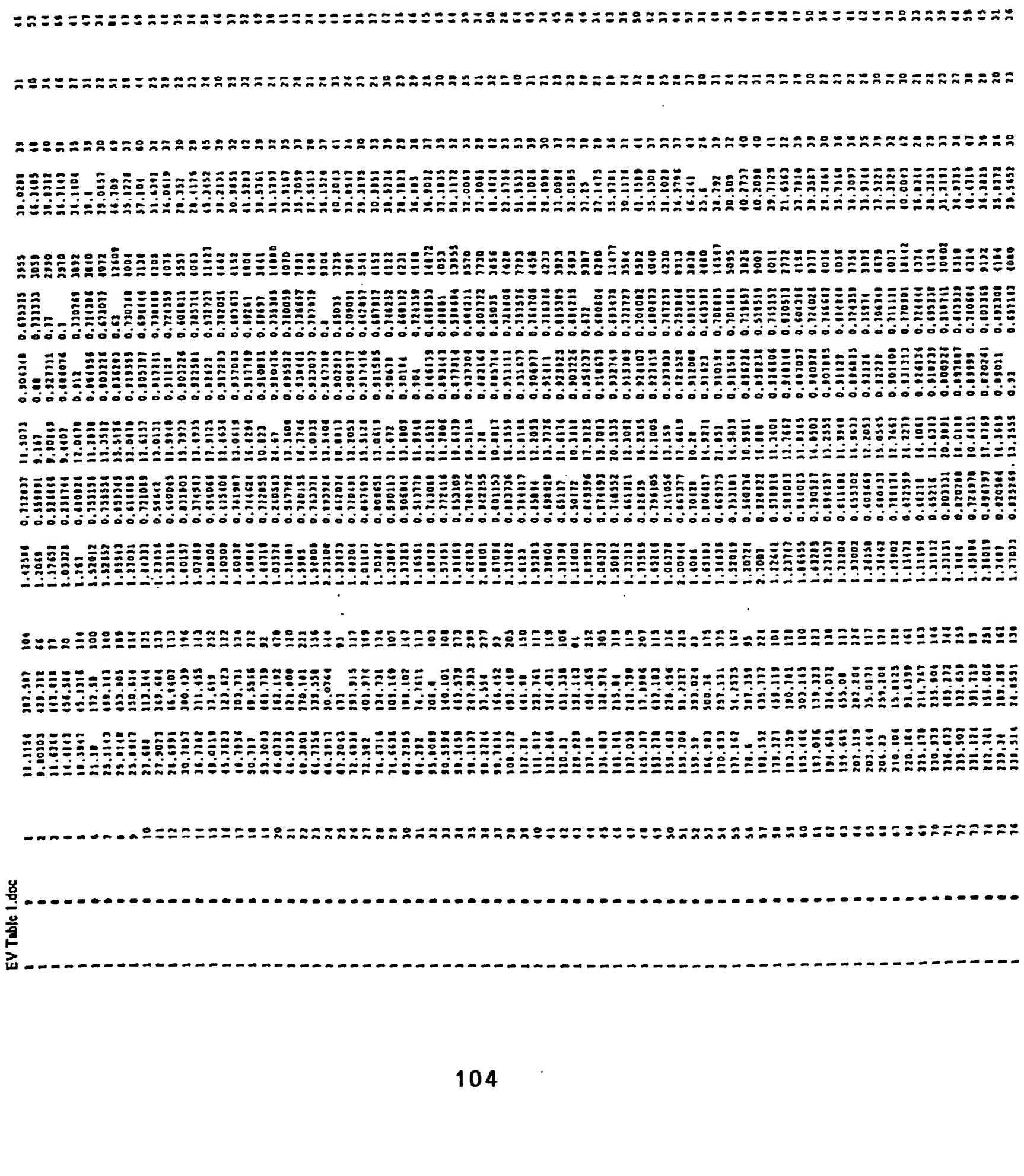
EV Table I.doc

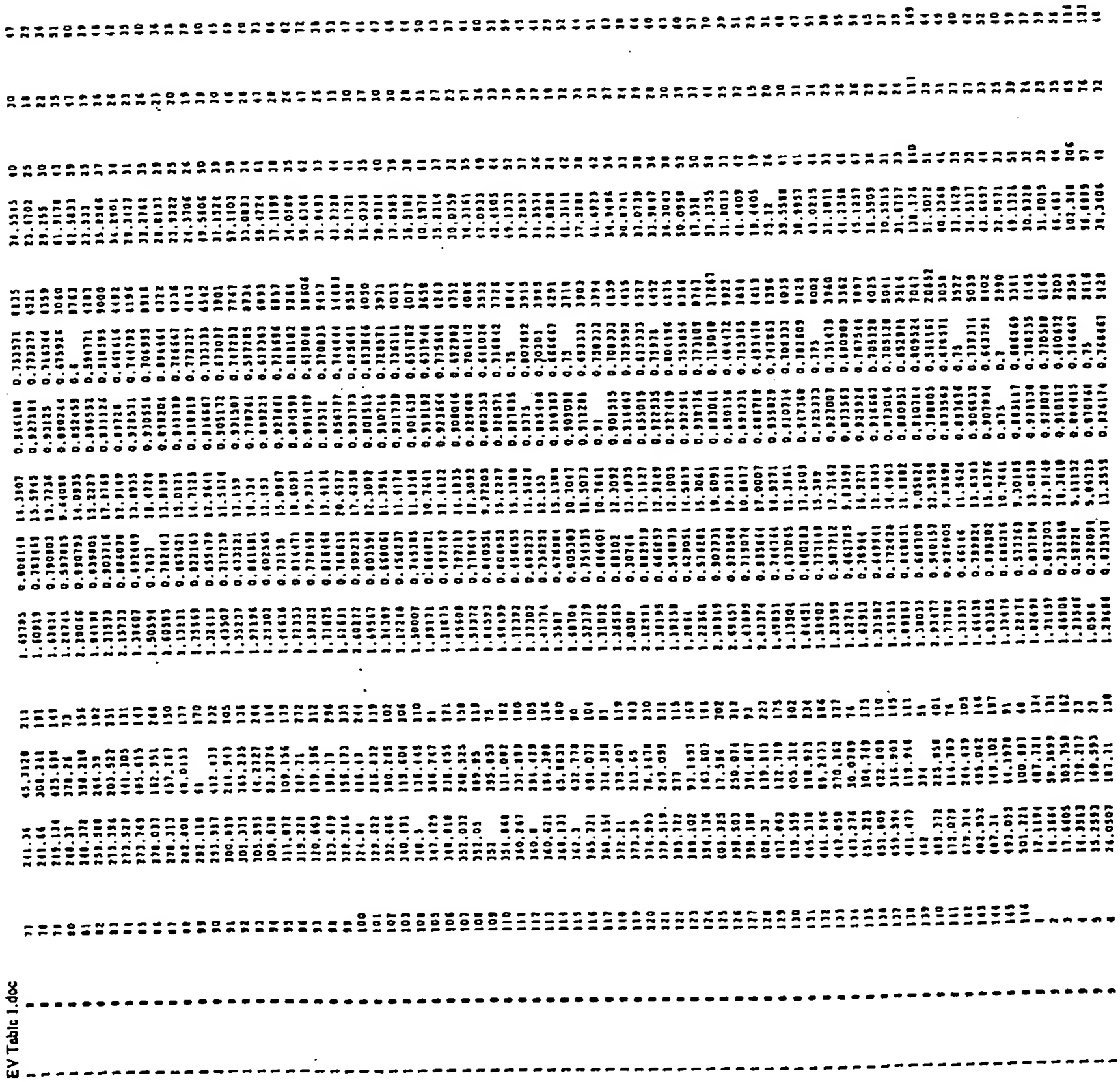


EV Table I.doc









EV Table 1.doc

1	9	83	283.646	101.811	127	1.6086	0.805787	12.7162	0.913669	0.661458	4625	36.4173	37	29	44
1	9	84	284.665	410.671	173	1.53054	0.759964	14.9415	0.925134	0.758772	7670	44.3353	46	35	54
1	9	85	281.69	289.917	84	1.53454	0.758511	10.3418	0.903226	0.666154	4211	50.131	51	39	62
1	9	86	286.312	80.4481	154	1.13337	0.470816	14.0028	0.916667	0.733333	4526	29.3896	30	23	36
1	9	87	292.589	305.311	90	1.08788	0.393714	10.7047	0.909091	0.763802	4134	45.9333	46	36	56
1	9	88	300.374	47.9346	107	1.37271	0.685059	11.472	0.89916	0.764286	4067	38.0093	39	29	46
1	9	89	302.5	332.755	106	1.53544	0.758039	11.6174	0.938053	0.84127	4700	44.3396	44	36	54
1	9	90	310.561	348.784	148	2.18785	0.908085	13.7273	0.840909	0.528571	8947	60.4527	60	41	74
1	9	91	305.159	364.341	44	1.29753	0.837205	7.48482	0.846154	0.6875	3760	85.4545	88	65	108
1	9	92	313.183	388.246	224	2.05965	0.874094	16.888	0.914286	0.59893	9248	41.2768	43	32	50
1	9	93	312.337	242.253	359	2.13587	0.883624	21.3797	0.810384	0.569841	19817	55.2006	58	42	89
1	9	94	315.082	322.148	382	1.48817	0.738934	15.2227	0.920571	0.736842	9474	52.0549	55	41	64
1	9	95	322.85	51.4724	143	1.82189	0.835902	14.4062	0.910615	0.839216	4364	26.773	28	20	32
1	9	96	327.5	256.405	158	1.28078	0.824816	14.1835	0.923977	0.752381	4988	31.5696	33	25	38
1	9	97	343.743	422.568	241	1.64439	0.793039	17.5172	0.895911	0.634211	9916	41.1452	42	31	51
1	9	98	342.921	48.3444	151	1.33418	0.661977	13.8658	0.920732	0.725962	4249	28.1391	29	22	34
1	9	99	344.143	139.607	26	1.12468	0.460666	5.97062	0.875	0.666667	5659	202.107	215	166	238
1	9	100	352.064	283.636	173	1.60986	0.783613	14.8415	0.935135	0.738772	4693	27.1618	28	21	33
1	9	101	358.062	396.634	145	1.37417	0.685884	13.5875	0.929487	0.74359	8423	58.0897	60	47	69
1	9	102	358.12	38.9816	217	2.18678	0.899316	18.6221	0.90195	0.645833	4595	21.1751	21	17	26
1	9	103	358.408	69.4079	76	1.72377	0.516431	9.83698	0.984118	0.690909	7203	94.8026	99	76	114
1	9	104	358.364	137.435	11	1.13876	0.478395	3.71241	0.785714	0.6875	210	19.0909	19	14	22
1	9	105	370.203	343.176	148	1.83856	0.839147	13.7273	0.89693	0.621048	4306	29.0946	30	22	35
1	9	106	373.012	202.324	186	2.28347	0.89901	16.5381	0.922272	0.65873	4596	27.6667	29	21	34
1	9	107	377.777	170.057	404	1.31494	0.650698	22.6801	0.930876	0.708294	20716	51.2772	52	38	64
1	9	108	384.198	243.899	328	1.4475	0.794718	20.4358	0.931818	0.788462	12022	36.6524	37	27	44
1	9	109	375.638	462.116	69	1.52244	0.75403	9.37302	0.896104	0.477273	3544	51.3623	52	41	62
1	9	110	393.181	491.839	155	1.6498	0.800655	14.0482	0.922619	0.645833	5724	36.9419	38	29	45
1	9	111	392.202	297.197	213	1.27925	0.623648	16.4682	0.924007	0.739583	6337	29.7512	30	22	38
1	9	112	391.713	407.231	395	1.80111	0.831707	15.757	0.924173	0.714286	9007	46.1893	48	35	58
1	9	113	404.037	455.366	82	1.57912	0.773937	10.2179	0.911111	0.759259	3945	48.1098	51	38	59
1	9	114	418.117	478.087	60	1.24455	0.595299	6.74039	0.669565	0.666667	4998	83.3	85	63	101
1	9	115	430.121	389.29	124	1.23756	0.599125	12.5851	0.905109	0.738095	8216	36	35	26	41
1	9	116	434.472	217.356	250	1.45536	0.72655	17.8412	0.91985	0.738994	8416	37.664	39	30	46
1	9	117	430.222	486.361	72	1.21482	0.567797	9.57481	0.878049	0.72	3740	51.9444	52	41	81
1	9	118	438.356	245.69	323	3.93418	0.967156	20.2195	0.813602	0.419101	11118	34.4241	36	27	42
1	9	119	444.212	381.171	316	2.27398	0.898116	20.0585	0.863388	0.548611	13939	56.169	58	43	71
1	9	120	441.239	288.096	230	1.75028	0.820708	17.1127	0.927418	0.684524	9481	41.1348	42	31	51
1	9	121	442.21	97.6643	143	1.34234	0.667091	13.4935	0.927501	0.744792	9280	57.3622	80	45	70
1	9	122	441.895	437.769	52	1.40181	0.700786	8.13696	0.996552	0.742857	3533	67.9423	70	53	81
1	9	123	450.907	139.883	181	2.22932	0.893749	14.3175	0.92	0.731818	4391	27.2133	27	20	35
1	9	124	452.118	424.193	119	1.29823	0.637708	12.3092	0.901515	0.708333	4146	34.8403	36	27	42
1	9	125	452.477	482.156	128	1.1575	0.503605	12.7662	0.920063	0.757396	4293	33.5391	35	26	40
1	9	126	455.773	440.5	44	1.23615	0.50786	7.48482	0.88	0.705714	3328	75.6364	77	62	90
1	9	127	474.277	407.954	130	1.41995	0.709952	12.8455	0.915493	0.714286	7021	54.0077	56	44	66
1	9	128	479.27	187.428	152	1.48858	0.740748	13.9116	0.853933	0.633333	4951	32.5724	33	26	40
1	9	129	479.157	215.288	121	1.27369	0.61919	12.4122	0.902985	0.715976	4343	35.8928	37	28	43
1	9	130	488.898	30.1837	69	1.4708	0.733302	7.89865	0.907407	0.777778	5891	120.224	122	92	147
1	9	131	485.018	132.565	188	1.6403	0.792675	14.6255	0.913333	0.777778	4290	25.5357	26	20	32
1	9	132	488.025	230.544	160	1.36575	0.68109	14.273	0.946748	0.784314	8081	50.5063	53	41	62
1	9	133	489.527	41.2515	183	1.47325	0.73435	14.4963	0.911602	0.705128	4206	25.4909	27	20	33
1	9	134	488.927	311.45	191	1.39813	0.697829	15.5945	0.927184	0.702206	9350	48.9529	50	38	60
1	9	135	493.98	94.3349	212	2.14011	0.885034	16.4294	0.921138	0.658185	7500	35.3774	38	29	43
1	9	136	495.135	339.708	192	1.32749	0.657471	15.6353	0.932039	0.761905	8831	45.9948	48	36	55
1	9	137	497.039	368.388	178	2.12112	0.881898	15.0365	0.93228	0.706349	6413	36.0281	38	27	44

Example of the summary output of AnalyseDNA.m program (summary for 10 3 by 3 montage images)

1	1187	163.912	79.3918	1.58849	0.398735	0.726461	0.137996	14.0612	3.315	0.905327	0.0350365	0.701248	0.0751176	6449.26	3496.95	41.539	18.352	42.9414	19.9393
	32.6103	14.336	50.5906	22.6586														
2	1305	169.036	86.8722	1.60511	0.399942	0.727835	0.138462	14.2634	3.43288	0.906571	0.0361453	0.70177	0.0730289	6786.37	3895.1	42.2416	17.0965	43.8243	17.343
	33.0245	13.3362	51.3609	23.2507														
3	1399	164.57	80.0679	1.58728	0.397935	0.72552	0.13163	14.0853	3.33904	0.905267	0.0361822	0.702891	0.0720005	6881.08	3525.81	44.7167	20.5918	46.3838	21.2267
	36.8313	16.2062	54.4617	29.1765														
4	1388	173.075	89.7145	1.60189	0.389682	0.727142	0.142518	14.3531	3.61912	0.902766	0.0374254	0.695945	0.0753089	6997.68	4212.87	43.1738	19.8582	44.9561	20.5112
	33.7348	15.9902	52.505	26.193														
5	1448	171.921	90.64	1.58867	0.400493	0.721084	0.141204	14.338	3.65065	0.906152	0.0379804	0.70023	0.0756884	7050.23	4163.04	44.0559	21.4761	45.8522	22.1282
	34.3315	16.6358	53.6367	26.5398														
6	1418	169.142	84.6806	1.60562	0.425512	0.728414	0.133721	14.0974	3.39686	0.904614	0.0362731	0.696204	0.0782855	6843.2	3924.12	44.266	19.3396	45.9485	19.8324
	34.5448	15.2864	53.8055	23.8472														
7	1756	129.864	99.2039	1.57806	0.451022	0.696813	0.185667	11.728	5.27441	0.893311	0.0481729	0.704526	0.0892393	5162.51	4393.11	34.9743	21.1969	36.16	22.0024
	28.8301	17.0421	43.0689	25.6082														
8	1280	171.387	84.1259	1.59301	0.405667	0.723421	0.137599	14.3833	3.40593	0.906384	0.0353179	0.702576	0.0751602	6865.87	3767.48	42.2752	17.2223	43.9781	17.7421
	33.0602	13.7194	51.3961	20.936														
9	1210	166.53	86.5094	1.60361	0.404986	0.72668	0.138568	14.1402	3.47784	0.905208	0.0356298	0.700331	0.0766773	6576.56	4023.38	41.6066	17.0994	43.163	18.2499
	33.3669	13.8149	50.6976	22.2407														
10	1425	159.606	83.618	1.57286	0.400372	0.717147	0.139085	13.8205	3.49536	0.904275	0.0379111	0.702759	0.07572	6507.18	3788.39	44.2141	20.7694	45.8765	21.136
	34.4582	16.3189	53.7958	28.6094														

WO 00/70528
PCT/US00/13154

CLAIMS

What is claimed is:

1. A method of predicting a property of a manipulation of cells based upon a descriptor, said method comprising:

providing a plurality of cells;

manipulating said plurality of cells;

capturing a morphological value from said plurality of cells;

assigning a degree of presence of said morphological value; and

storing said morphological value and said degree of presence;

wherein said descriptor is derived from a first component of a cell and a second component of said cell, said capturing said morphometric value from said plurality of cells comprises determining a relationship between said first component and said second component.

2. The method of claim 1 wherein said first component and said second component are selected from a protein, a protein modification, a nucleic acid, a lipid, a carbohydrate, a subcelluar structure and an organelle.

3. The method of 1 wherein said step of manipulation occurs in a manner selected from a electrical source, a chemical source, a thermal source, a gravitational source, a nuclear source, a temporal source, and a biological source

4. The method of claim 3 wherein said chemical source is selected from a pharmacological candidate and a drug screening library.

5. The method of claim 1 wherein said morphological value is selected from a count, an area, a perimeter, a length, a breadth, a fiber length, a fiber breadth, a shape factor, a elliptical form factor, an inner radius, an outer radius, a mean radius, an equivalent radius, an equivalent sphere volume, an equivalent prolate volume, an equivalent oblate volume, an equivalent sphere surface area, an average gray value, a total gray value, and an optical density.

6. The method of claim 1 wherein said degree of presence is multiple of a quantized value.

7. A computer program product for populating a database with manipulated biological information, said computer program product comprising:

code for providing a plurality of cells in various stages of the cell cycle, said stages of the cell cycle including at least one selected from interphase, G0 phase, G1 phase, S phase, G2 phase, M phase, prophase, prometaphase, metaphase, anaphase, and telophase;

code for manipulating said cells in said various stages of cell cycle development to form a plurality of manipulated cells;

code for capturing an image of said plurality of manipulated cells;

code for determining a descriptor from said image for said manipulated cells;

code for populating a database with said descriptor;

wherein said image includes a first component of a cell and a second component of said cell; and

a computer readable storage medium for holding the codes.

8. The computer program product of claim 7 wherein said first component and said second component are selected from a protein, a protein modification, a nucleic acid, a lipid, a carbohydrate, a sub-cellular structure and an organelle.

9. The computer program product of claim 7 wherein said image is a digitized representation of said plurality of manipulated cells.

11. The computer program product of claim 9 wherein said digitized representation provides a density value of said plurality of manipulated cells.

11. The computer program product of claim 7 wherein said descriptors comprise numeric or logical values.

12. The computer program product of claim 11 wherein said values further comprises a nucleotide.

13. The computer program product of claim 11 wherein said values further comprises an amino acid letter.

14. A system for capturing images of cells or cell structures, the system comprising:

a cell holder comprising a plurality of sites in a spatial orientation, each of the sites being capable of holding a plurality of cells to be imaged;

an image capturing device coupled to the cell holder, the image capture device being adapted to capture at least one image in at least one of the plurality of sites;

an illumination apparatus comprising a liquid light guide coupled to the plate for highlighting the plurality of cells in a relatively even spatial manner for image capturing purposes;

an image processing device coupled to the image capturing device, the image capturing device being adapted to convert the image into a digital representation; and

a database storage device comprising a database management element coupled to the image capturing device, the database storage device being adapted to retrieve the digital representation of the image from the image processing device and storing the digital representation.

15. The system of claim 14 further comprising a stage comprising a device for moving the cell holder in a spatial direction to traverse across the cell holder in the spatial orientation.

16. The system of claim 14 wherein the illumination apparatus comprises sub-elements, at least one of the sub-elements being positioned away from the image capturing device to prevent a possibility of vibration from the one sub-elements to be transmitted to the image capturing device.

17. The system of claim 14 wherein the digital representation comprises a plurality of regions and objects.

18. The system of claim 14 further comprising a computing device connected between the database storage device and the image processing device.

19. The system of claim 14 wherein the image capturing device comprises a magnification of at least 1X and greater to capture the image of the site.

20. The system of claim 14 wherein the plurality of sites comprises at least 96 sites.

21. The system of claim 14 wherein the liquid light guide characterized as a flexible member that substantially prevents vibration from the an element of the illumination apparatus to be transferred to the image capturing device.

22. The system of claim 14 wherein the spatial direction can be selected from an x-direction, a y-direction, or a z-direction in a Cartesian coordinate system.

23. The system of claim 14 wherein the each of the sites comprises a volume that is sufficient to prevent a solution therein from evaporating in a substantial manner that may influence the image capturing.

24. A method for identifying a mechanism of action for a first compound, the method comprising the steps of:

receiving the first compound;

measuring at least one feature of a cellular phenotype to define a target phenotype;

identifying additional compounds providing a feature similar to the feature identified in the measuring step; and

characterizing the first compound in terms of distance from a specific target phenotype having known characteristics.

25. The method of claim 24 comprising the further step of storing the additional compounds and their associated phenotypes in a database.

26. A method for identifying a mechanism of action for a cellular stimulus, the method comprising the steps of:

receiving cells of interest;

measuring at least one feature of the cells to define and quantify a target phenotype;

identifying additional compounds providing a feature similar to the feature identified in the measuring step; and

characterizing the first compound in terms of distance from a specific target phenotype having known characteristics.

27. A method for identifying information relevant to at least one of a mechanism of action and cellular activity by utilizing assay data to elucidate a phenotype, the method comprising the steps of:

identifying a target protein;

identifying positive and negative biochemical hits related to the target protein;

defining the target phenotype utilizing the positive and negative hits;

identifying other compounds providing similar features; and

characterizing the first compound in terms of distance from a specific target phenotype having known characteristics.

5

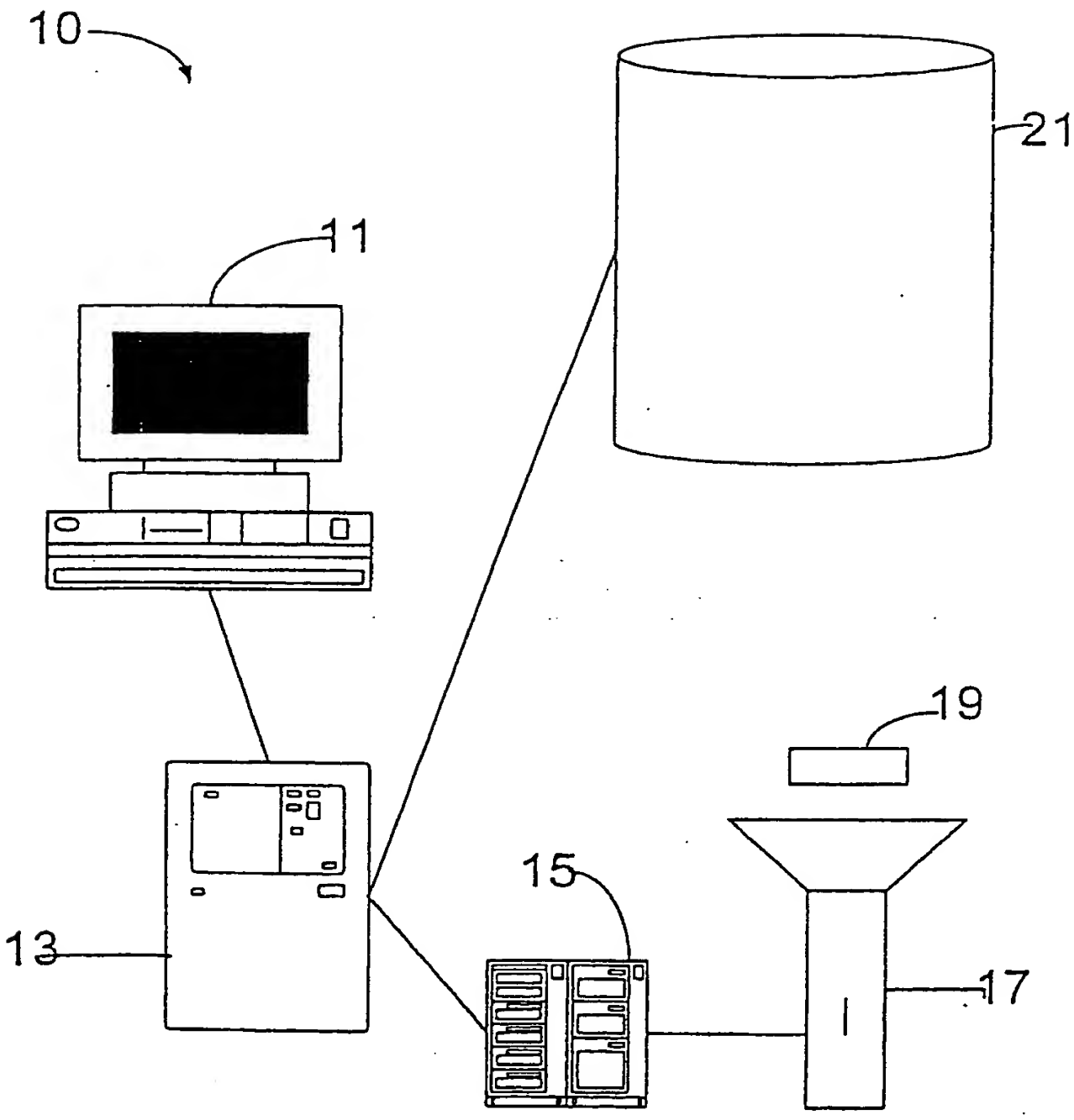


FIG. 1

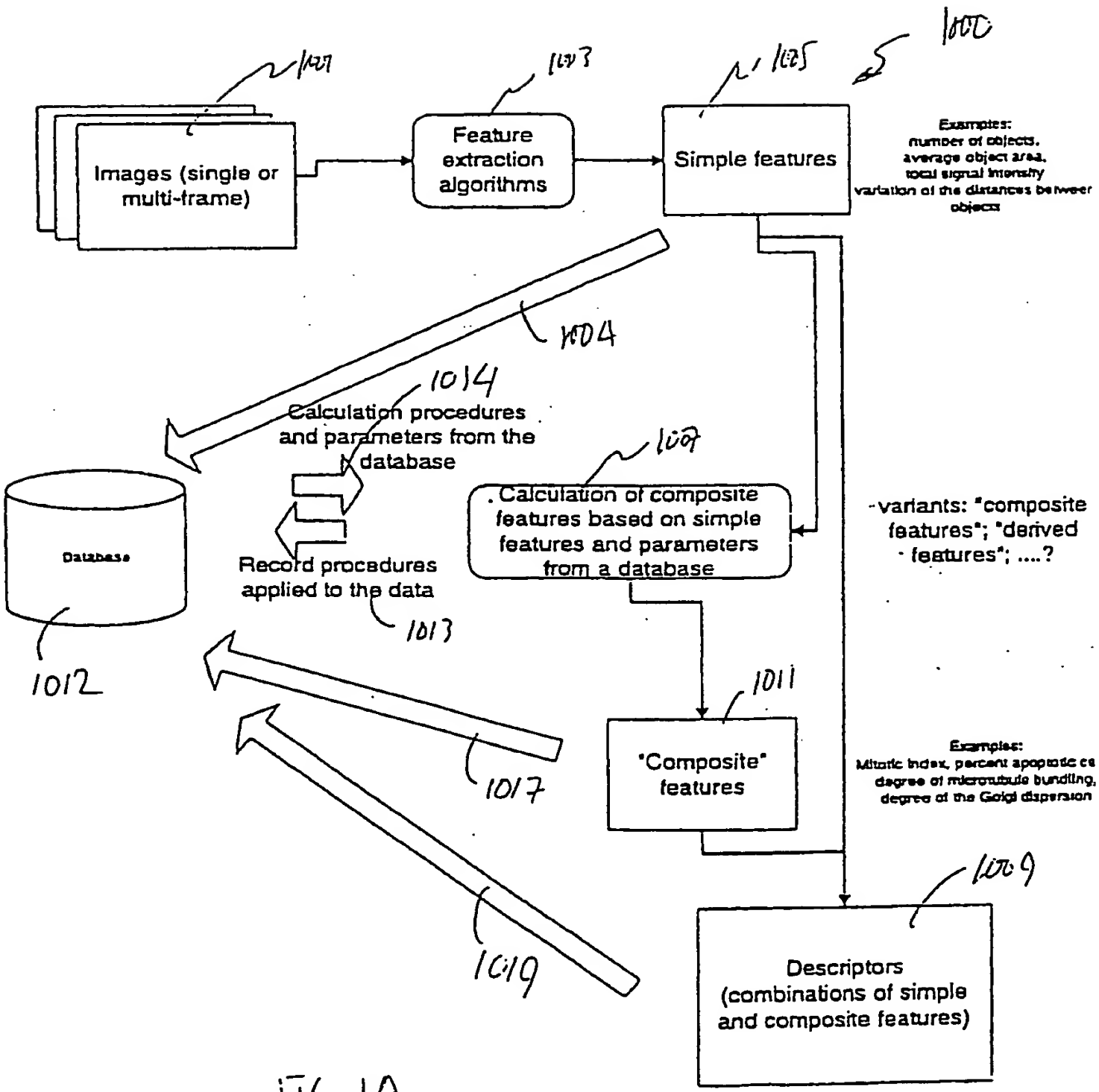

1000
1001
1003
1005
Images (single or multi-frame)
Feature extraction algorithms
Simple features
Examples:
number of objects,
average object area,
total signal intensity
variation of the distances between objects
1004
1014
Calculation procedures and parameters from the database
1007
Calculation of composite features based on simple features and parameters from a database
-variants: "composite features"; "derived features"; ....?
Database
Record procedures applied to the data
1013
1012
1011
1017
"Composite" features
Examples:
Mitotic index, percent apoptotic ce
degree of microtubule bundling,
degree of the Golgi dispersion
1019
1009
Descriptors (combinations of simple and composite features)

FIG. 1A

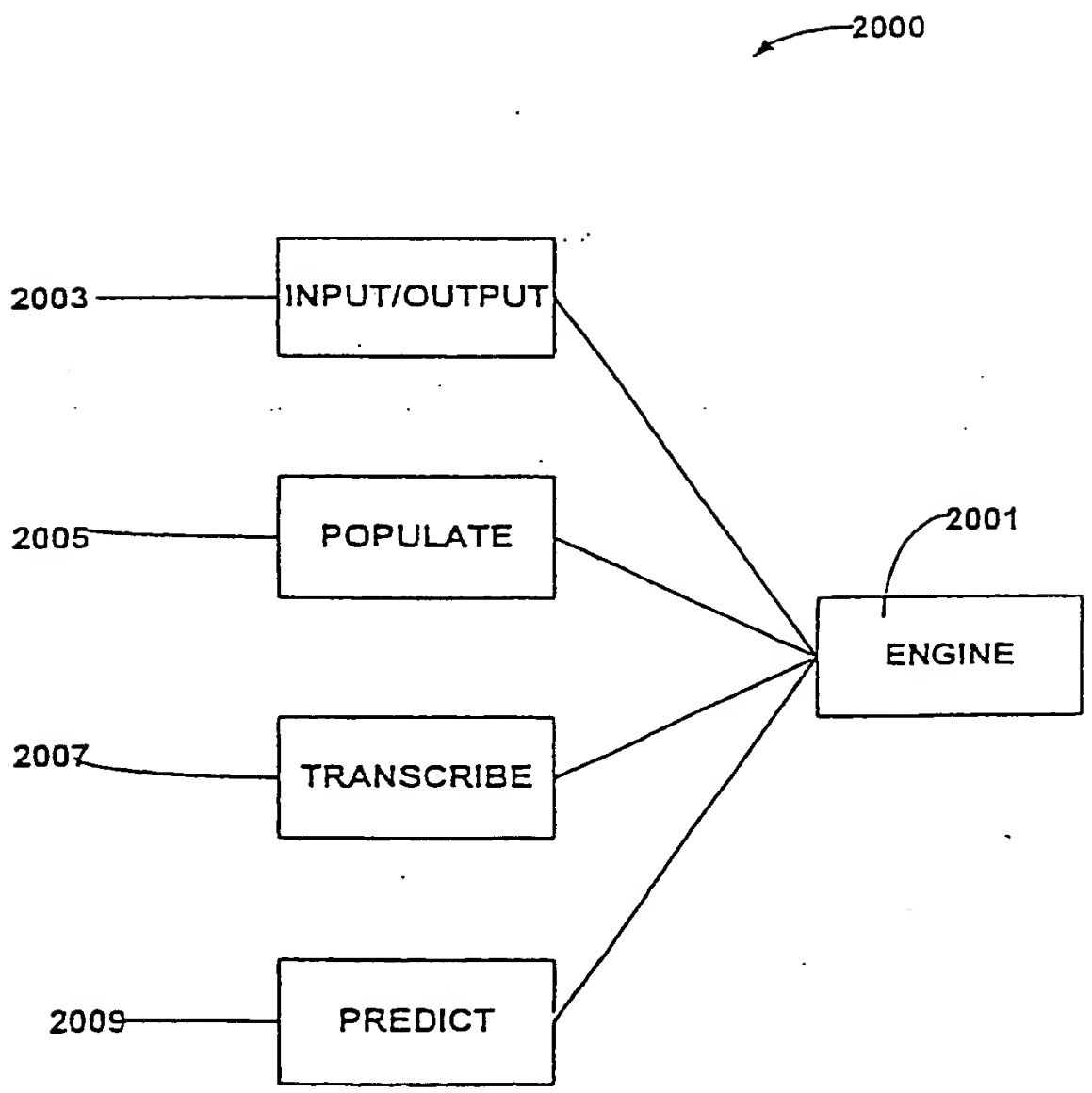

2000
2003
INPUT/OUTPUT
2005
POPULATE
2007
TRANSCRIBE
2009
PREDICT
2001
ENGINE

FIG. 1B

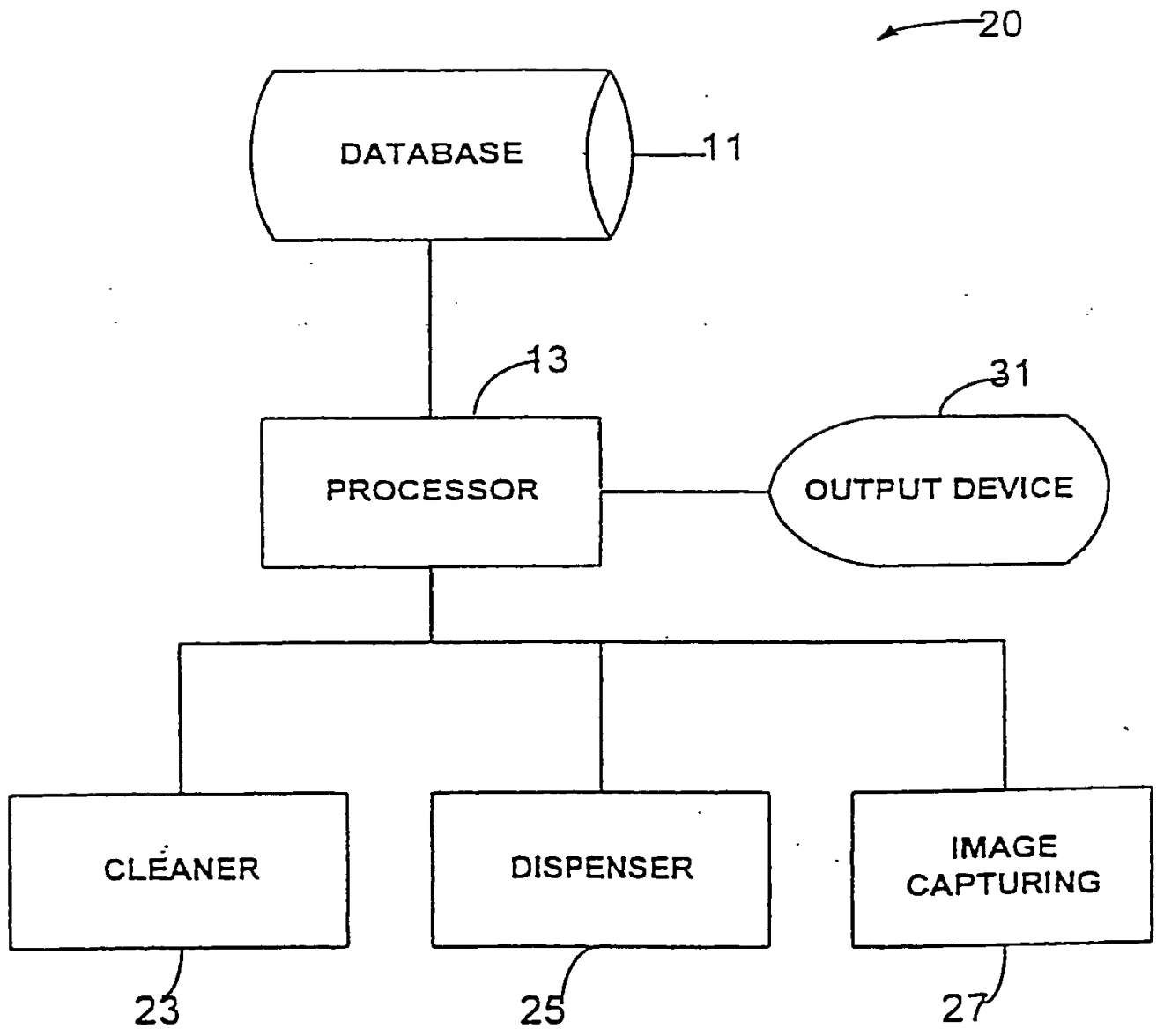

20
DATABASE
11
13
31
PROCESSOR
OUTPUT DEVICE
CLEANER
DISPENSER
IMAGE CAPTURING
23
25
27

FIG. 2

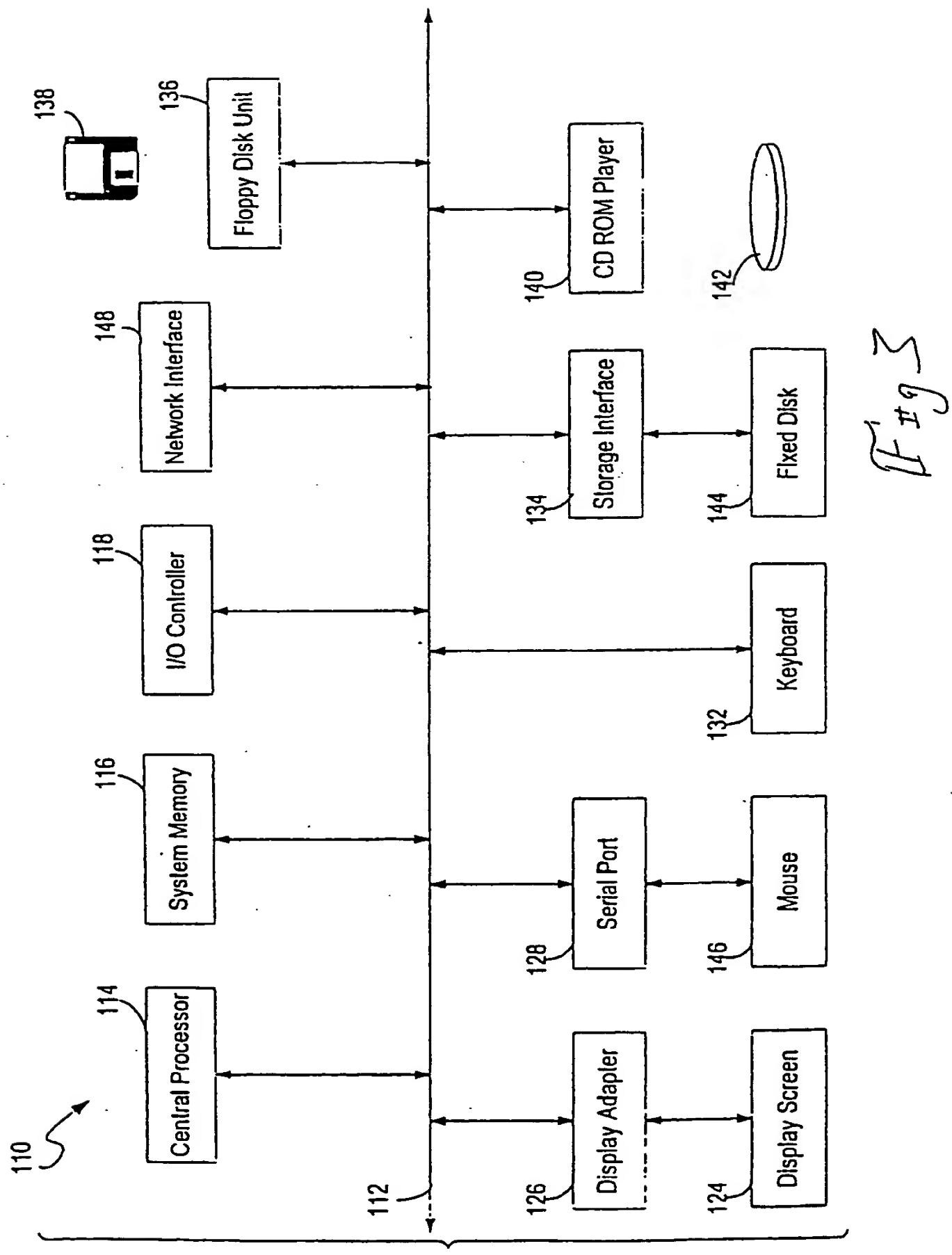

110
Central Processor
114
System Memory
116
I/O Conlroller
118
Network Interface
148
Floppy Disk Unit
136
138
CD ROM Player
140
142
Storage Interface
134
Fixed Disk
144
Keyboard
132
Serial Port
128
Mouse
146
Display Adapter
126
Display Screen
124
112

Fig 3

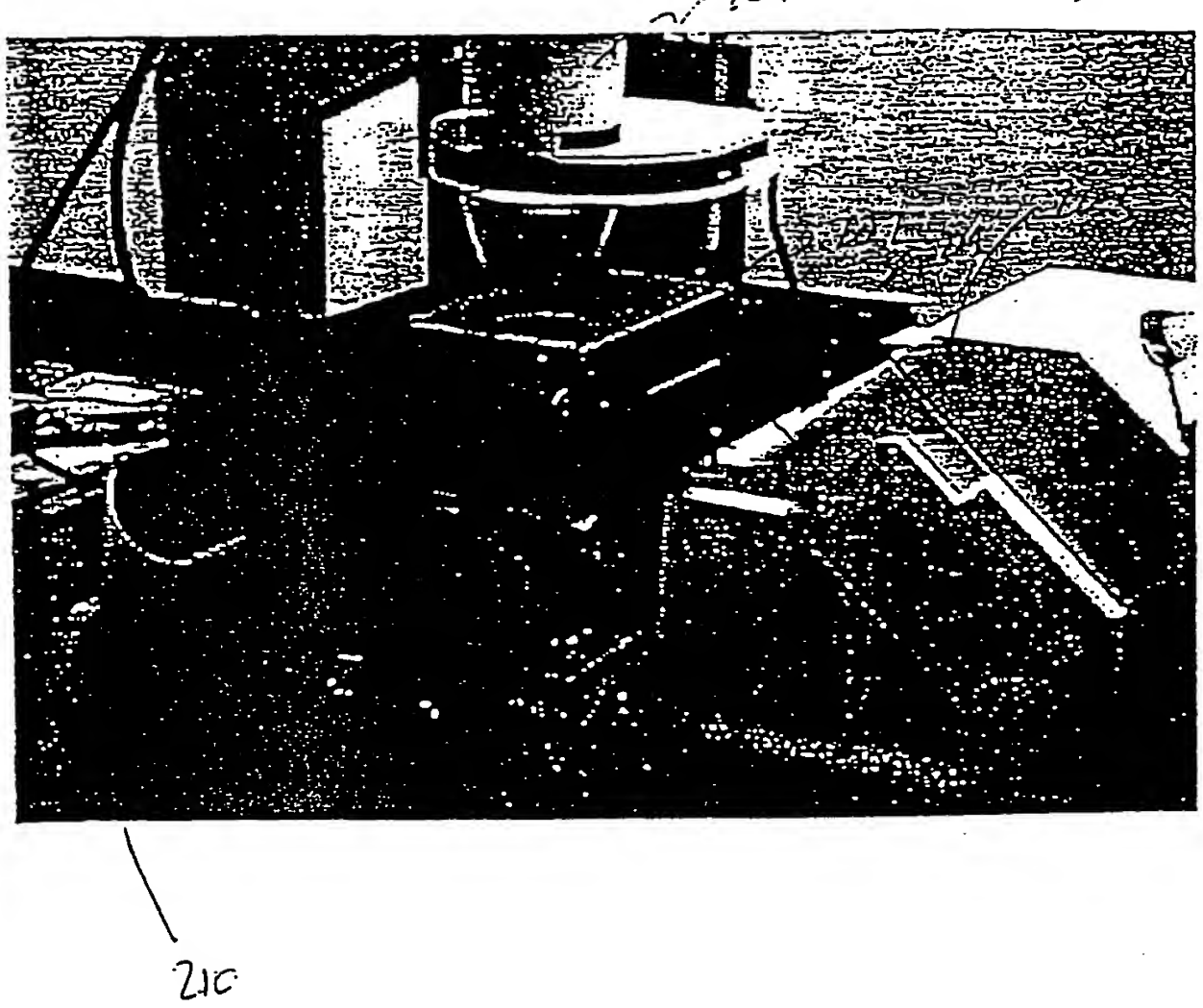


FIG. 4

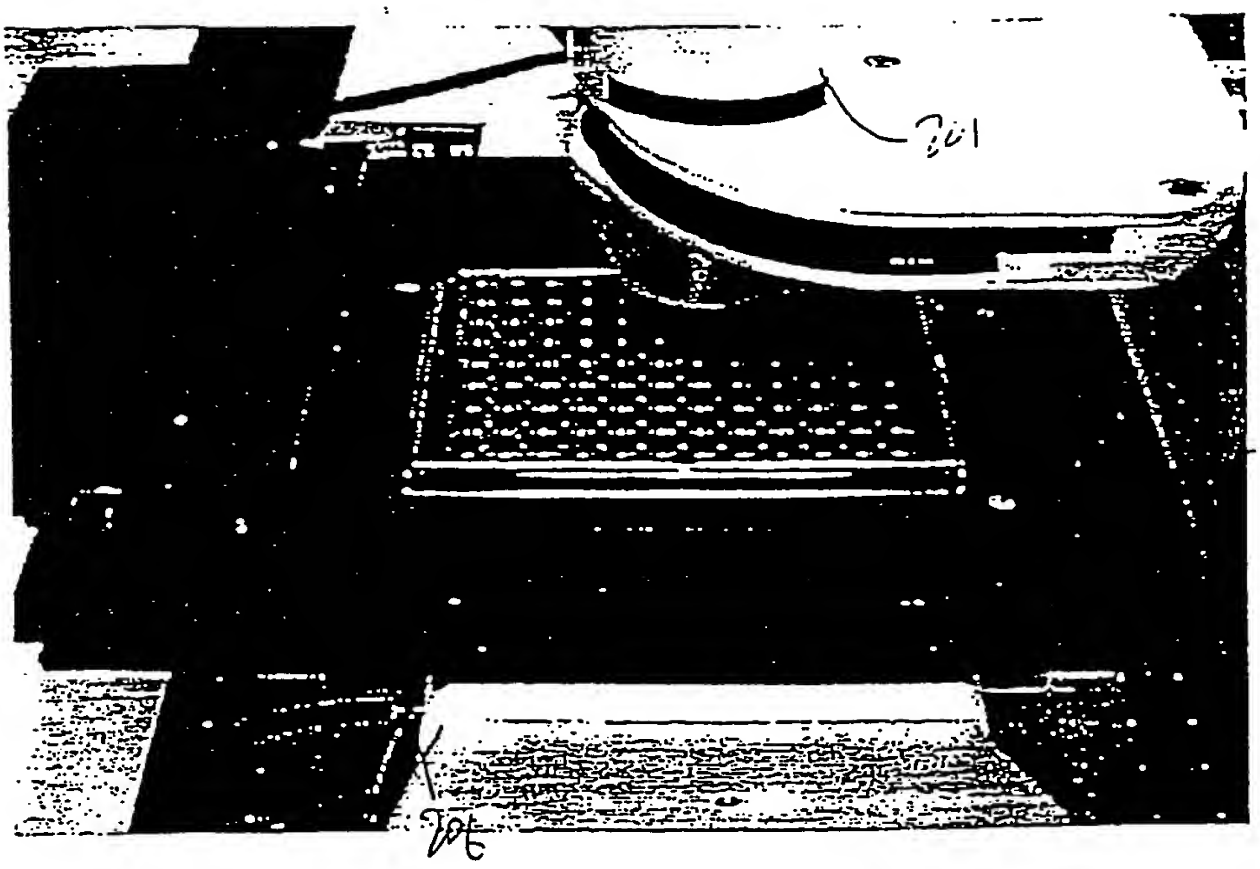


FIG 5

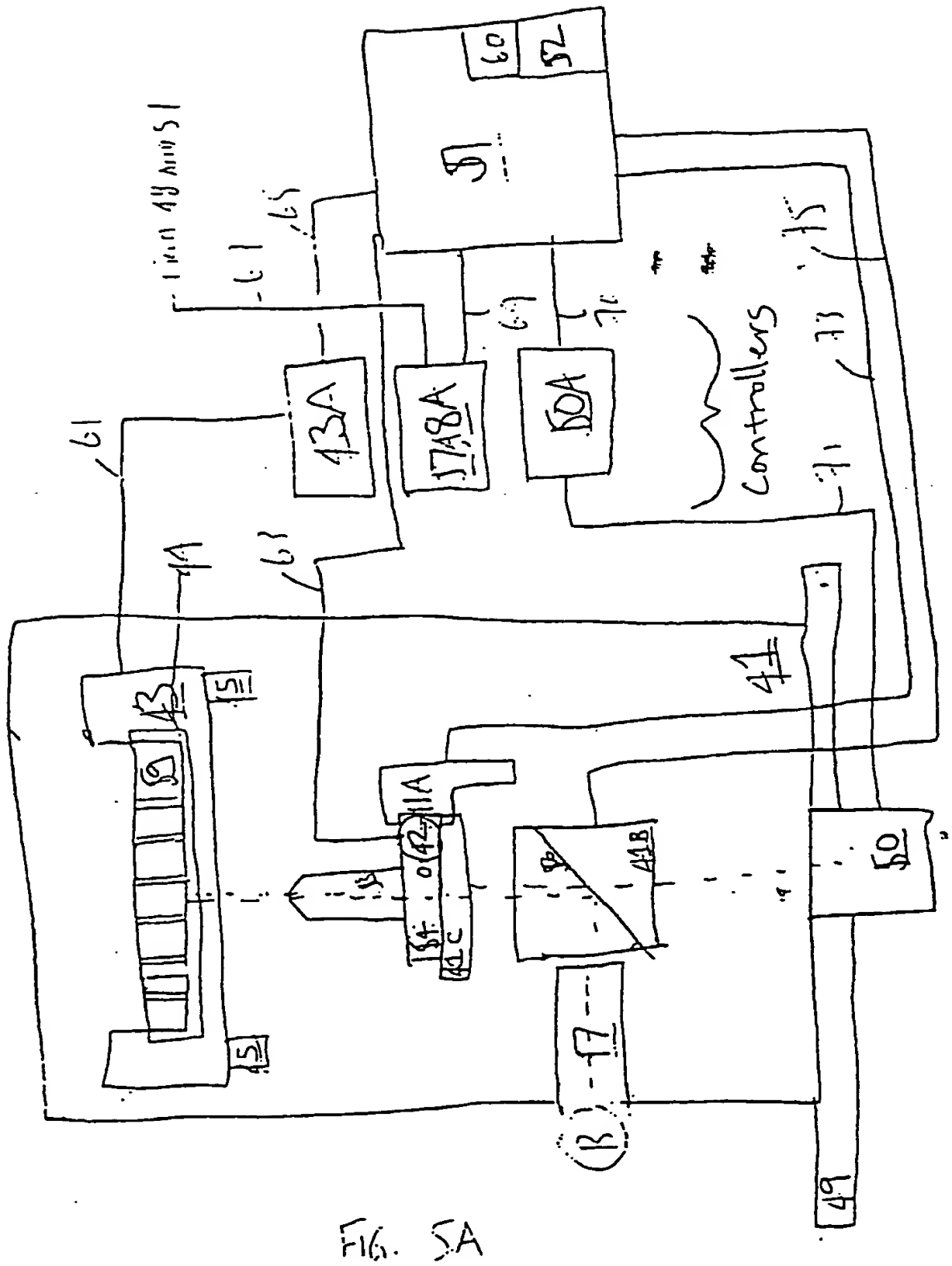

60
52
51
Controllers
43A
17A&B
50A
61
43
51
45
41
11A
42
44C
44B
50
17
49

FIG. 5A

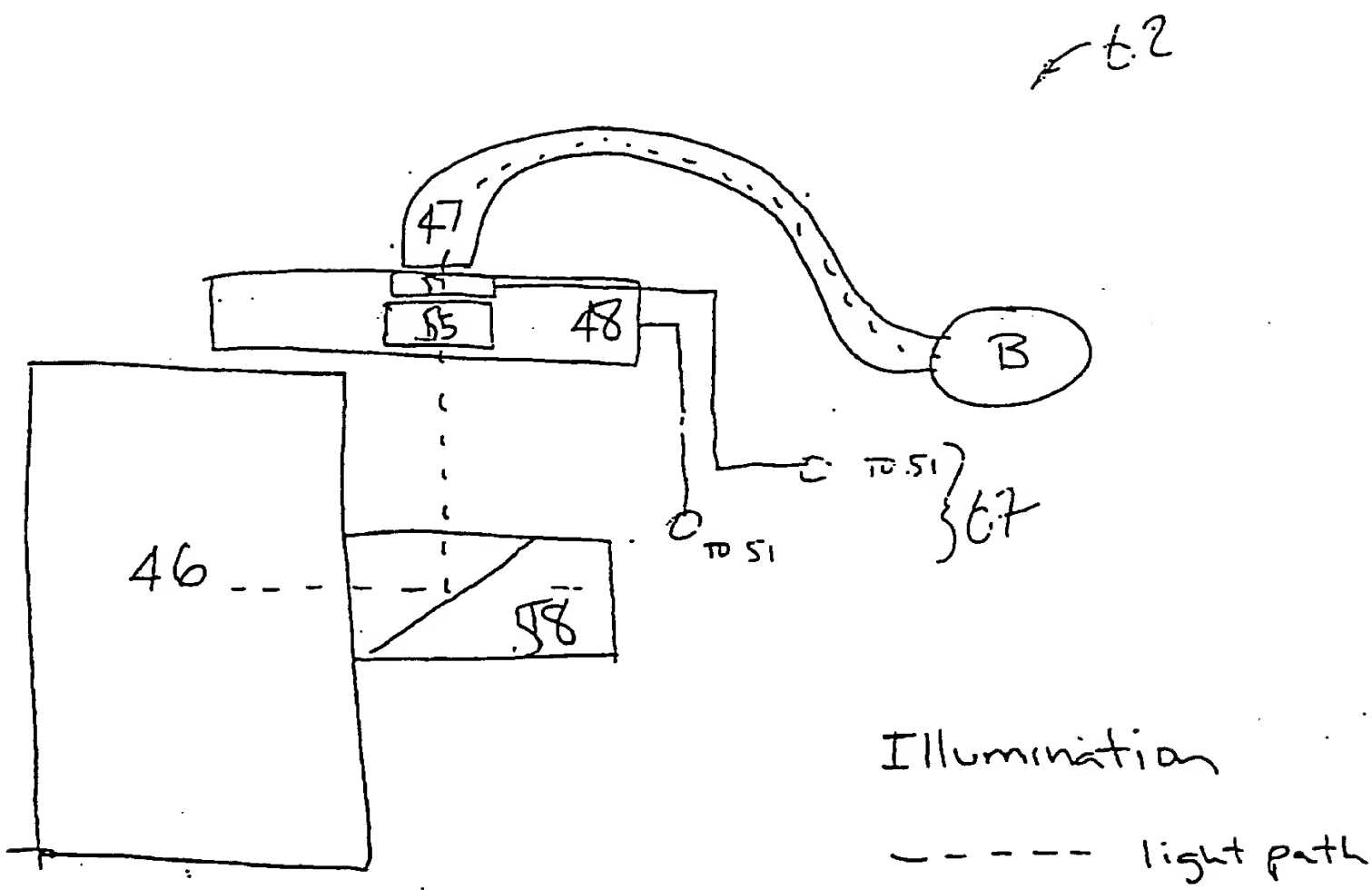


FIG. 5B

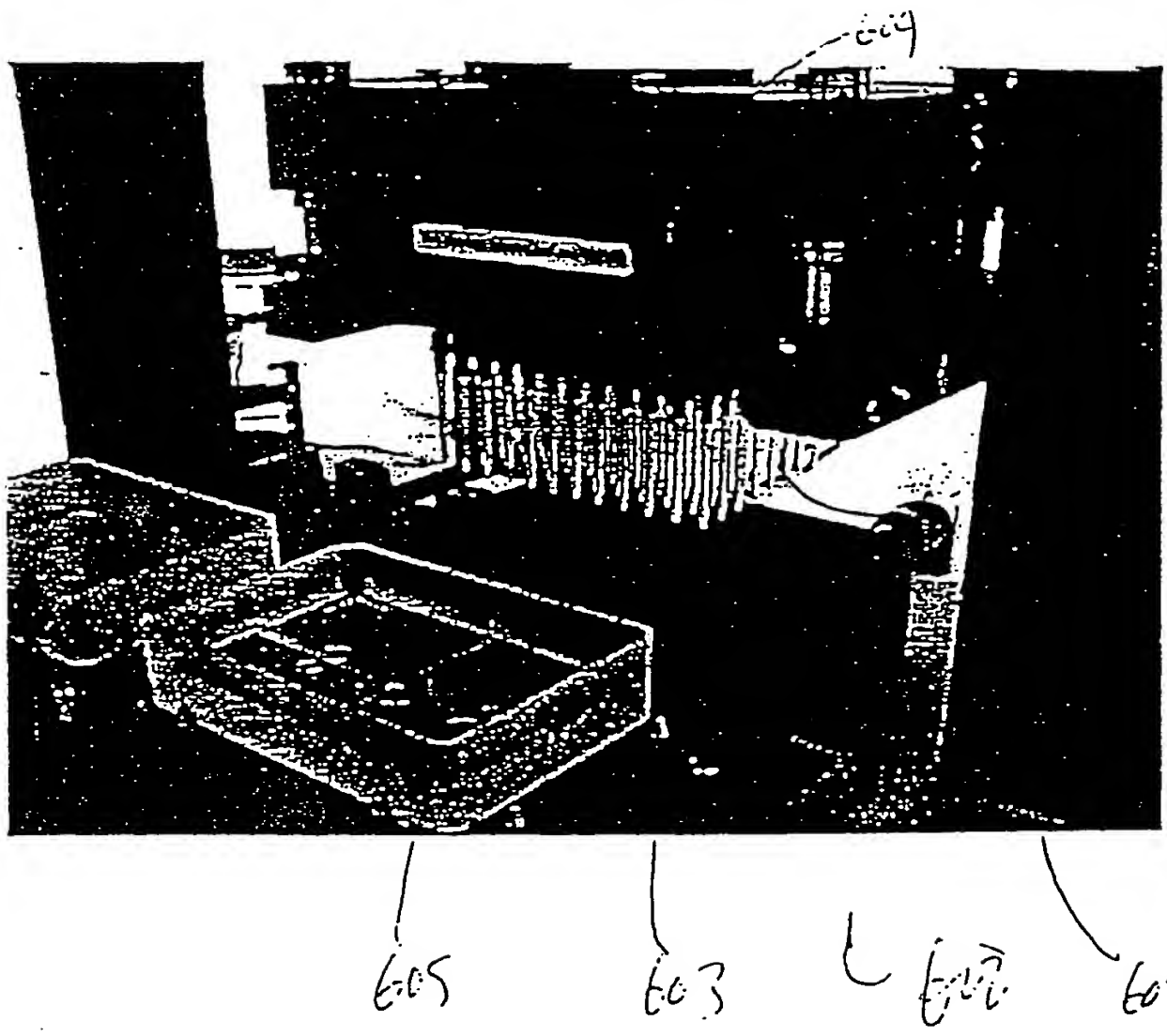


FIG. 6

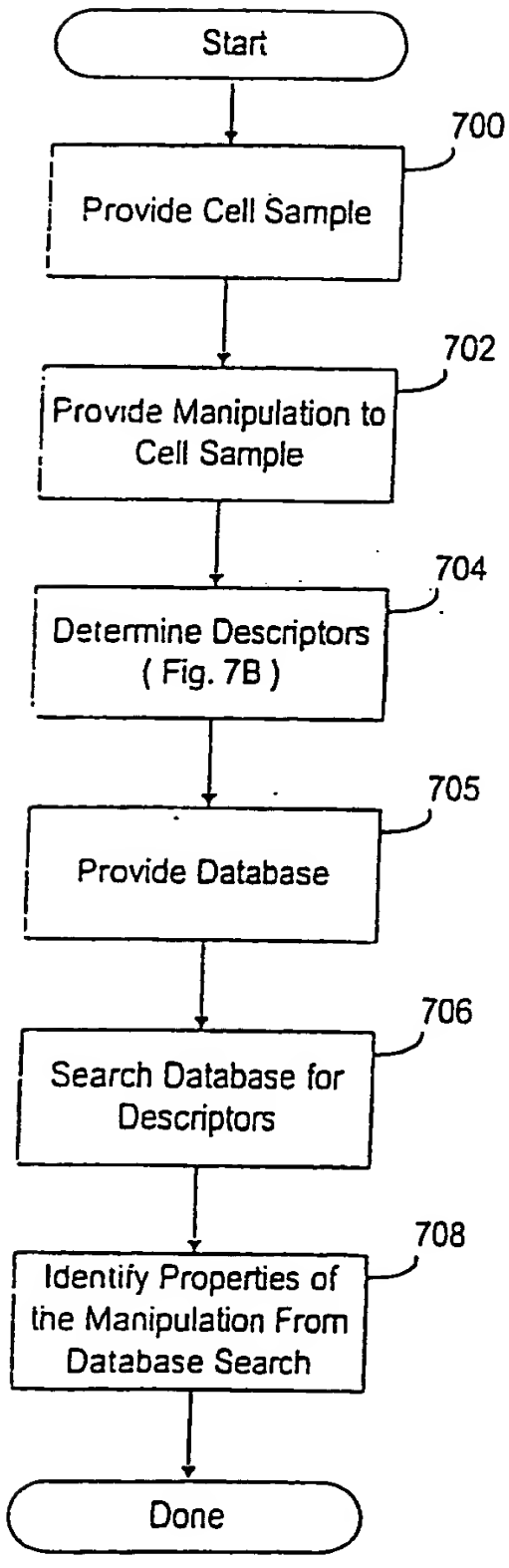

Start
700
Provide Cell Sample
702
Provide Manipulation to Cell Sample
704
Determine Descriptors (Fig. 7B)
705
Provide Database
706
Search Database for Descriptors
708
Identify Properties of the Manipulation From Database Search
Done

Fig. 7A

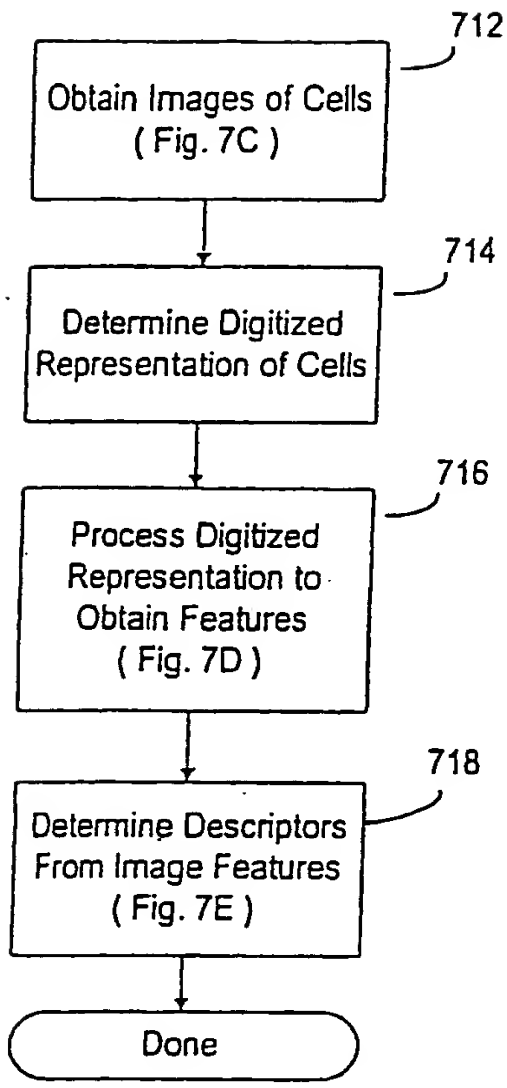


Fig. 7B
Step 704 of Fig. 7A

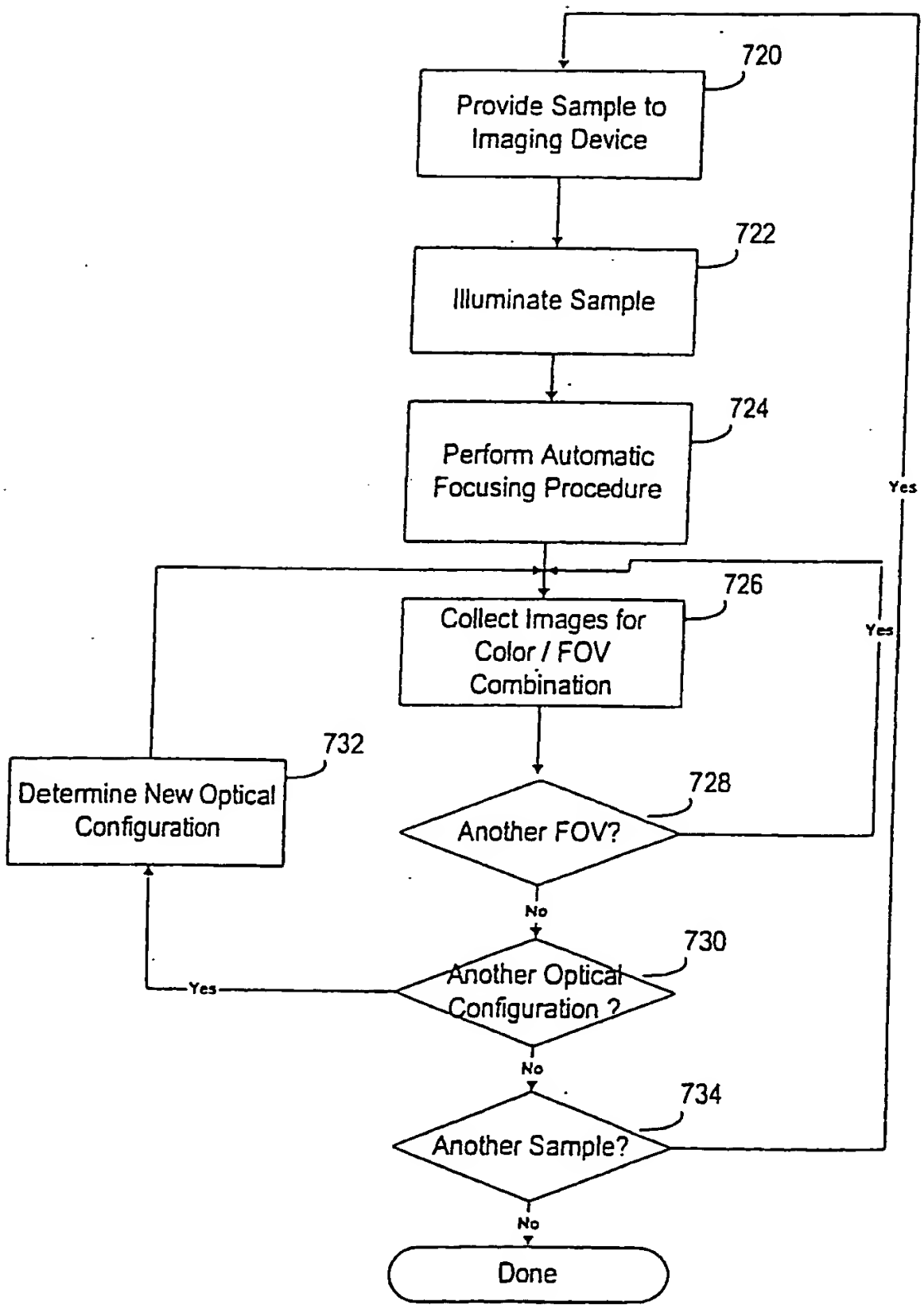


Fig. 7C
Step 714 of Fig. 7B

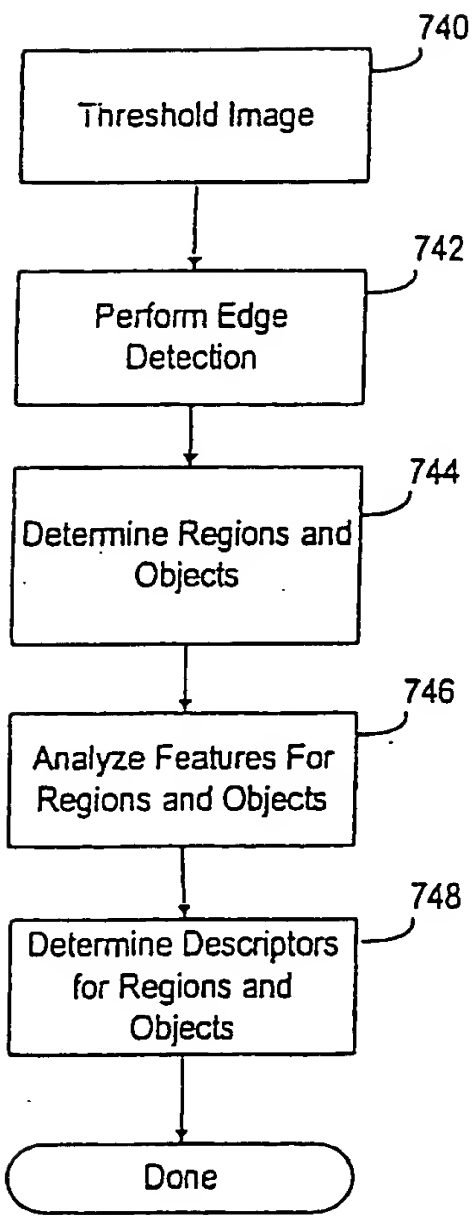


Fig. 7D
Step 716 of Fig. 7B

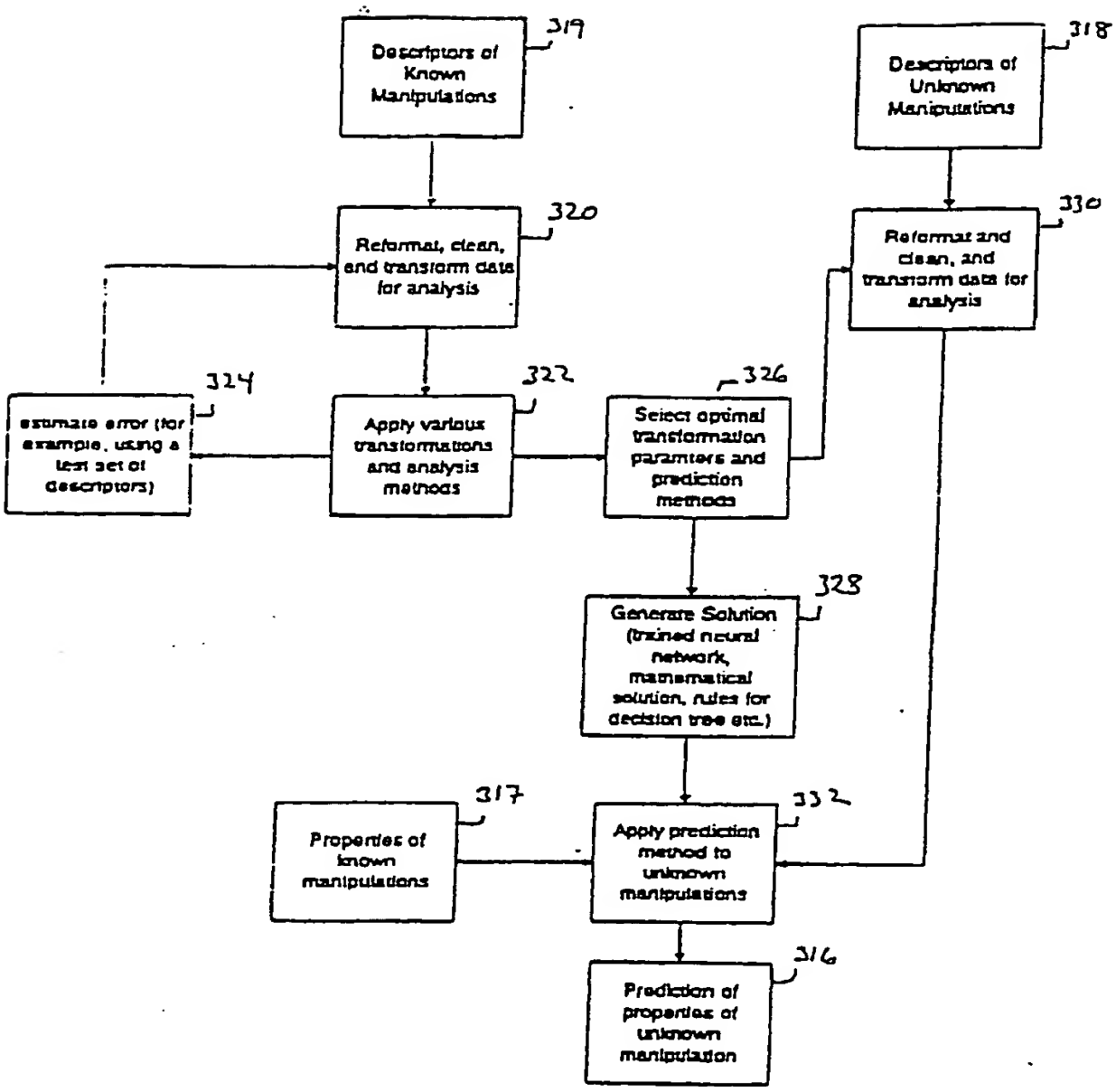

Descriptors of Known Manipulations
319
Descriptors of Unknown Manipulations
318
Reformat, clean, and transform data for analysis
320
Reformat and clean, and transform data for analysis
330
estimate error (for example, using a test set of descriptors)
324
Apply various transformations and analysis methods
322
Select optimal transformation parameters and prediction methods
326
Generate Solution (trained neural network, mathematical solution, rules for decision tree etc.)
328
Properties of known manipulations
317
Apply prediction method to unknown manipulations
332
Prediction of properties of unknown manipulation
316

FIG. 7E

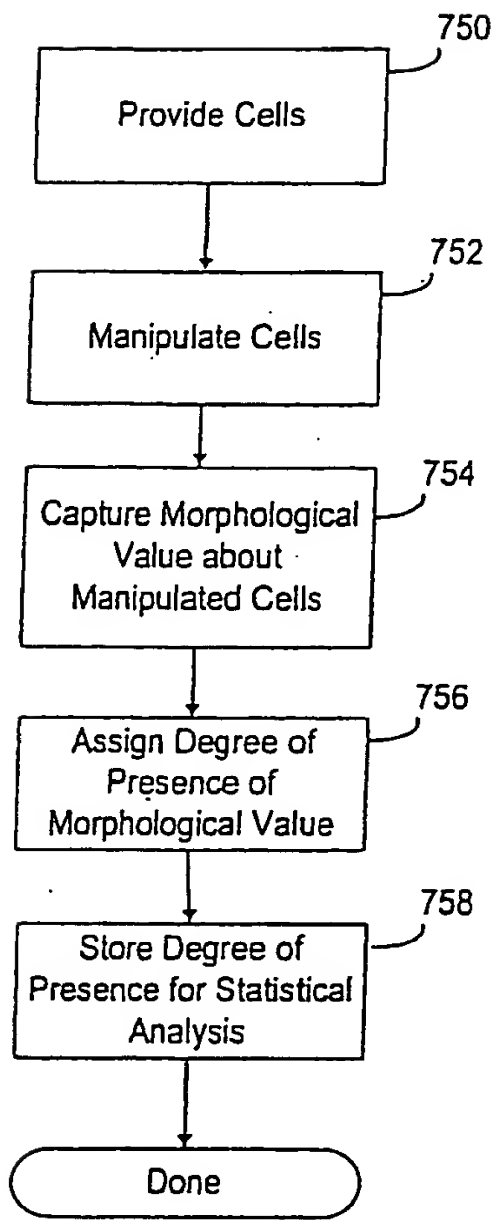

750
Provide Cells
752
Manipulate Cells
754
Capture Morphological Value about Manipulated Cells
756
Assign Degree of Presence of Morphological Value
758
Store Degree of Presence for Statistical Analysis
Done

Fig. 7F

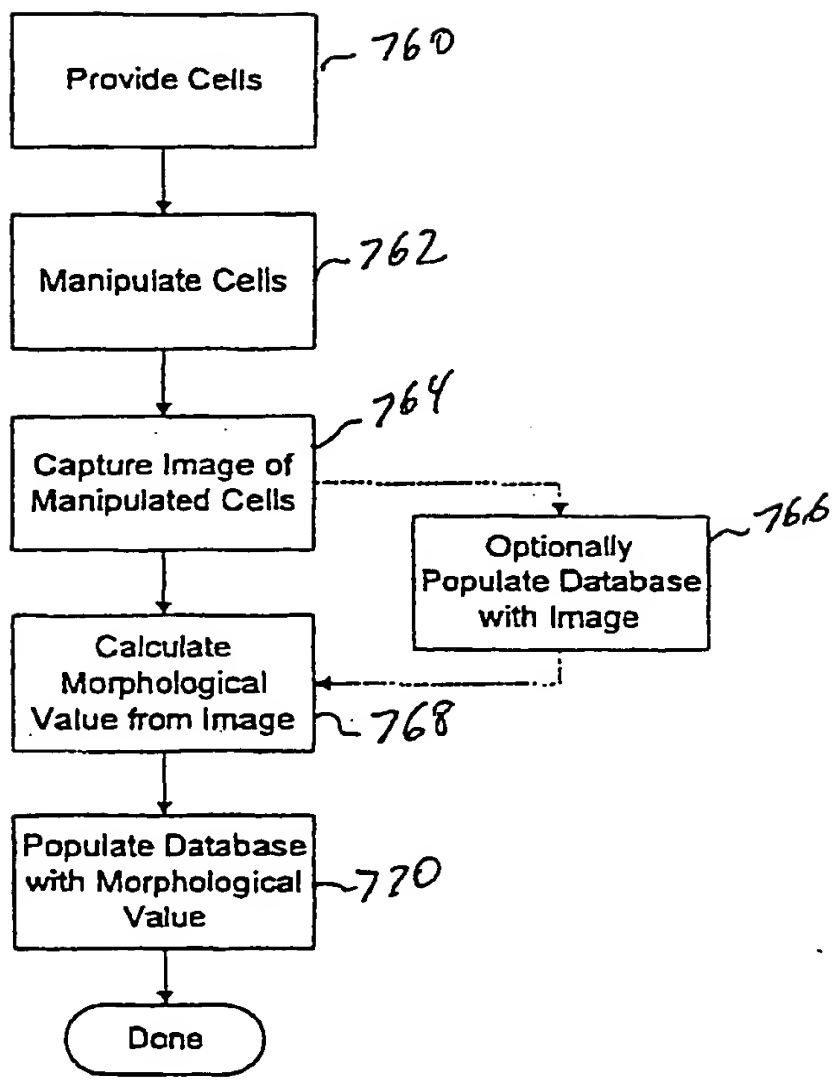

Provide Cells
760
Manipulate Cells
762
Capture Image of Manipulated Cells
764
Optionally Populate Database with Image
766
Calculate Morphological Value from Image
768
Populate Database with Morphological Value
770
Done

Fig 7G

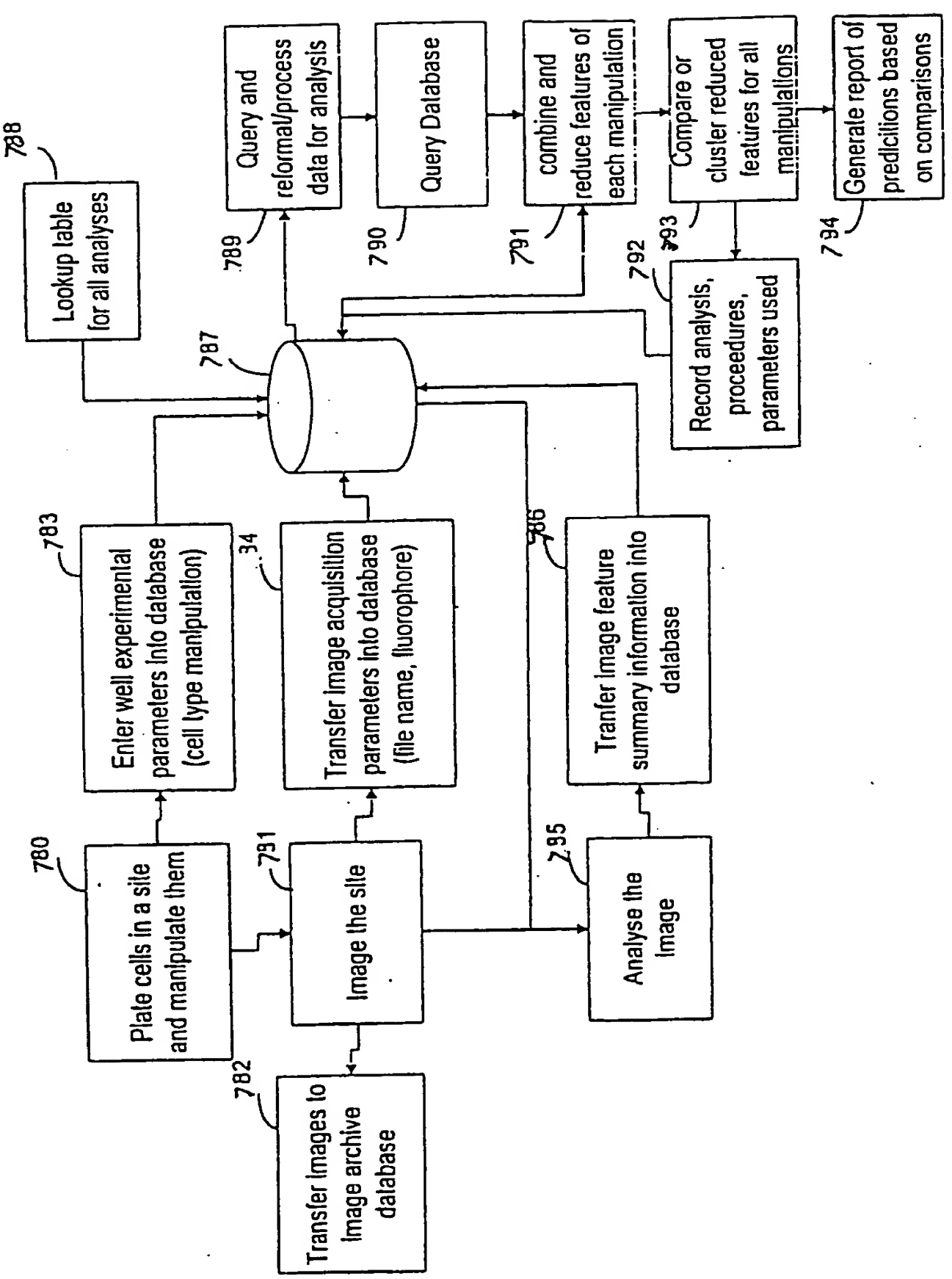

Lookup lable for all analyses
Query and reformat/process data for analysis
Query Dalabase
combine and reduce features of each manipulation
Compare or cluster reduced features for all manipulations
Generate report of predicitions based on comparisons
Record analysis, proceedures, parameters used
Enter well experimental parameters into database (cell type manipulation)
Transfer Image acquisition parameters into dalabase (file name, fluorophore)
Tranfer Image feature summary information into database
Plate cells in a site and manipulate them
Image the site
Analyse the Image
Transfer Images to Image archive dalabase
788
789
790
791
792
793
794
787
783
34
786
780
781
785
782

Fig.7H

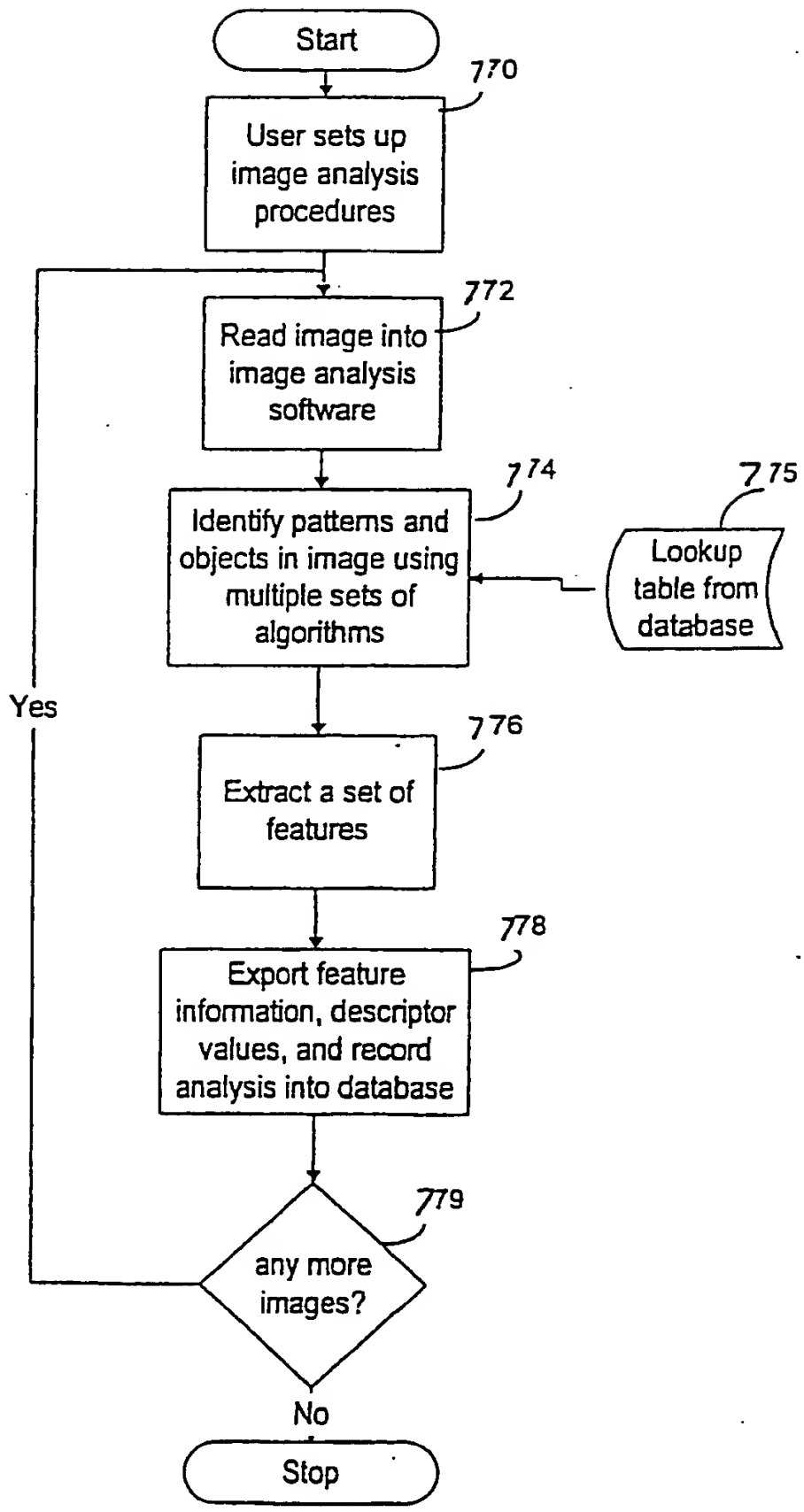

Start
770
User sets up image analysis procedures
772
Read image into image analysis software
774
Identify patterns and objects in image using multiple sets of algorithms
775
Lookup table from database
776
Extract a set of features
778
Export feature information, descriptor values, and record analysis into database
779
any more images?
Yes
No
Stop

Fig. 71

11/24

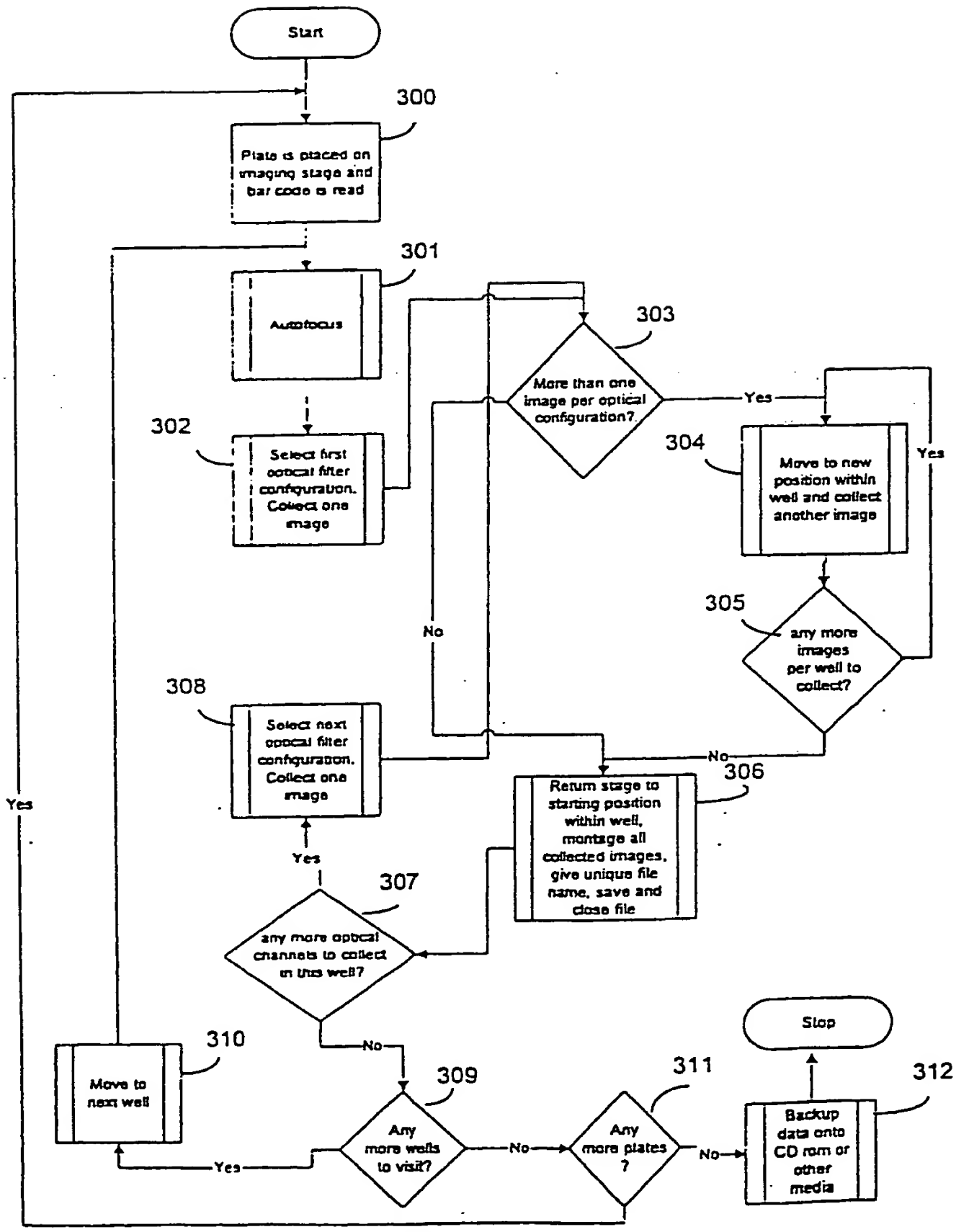

Start
300
Plate is placed on imaging stage and bar code is read
301
Autofocus
302
Select first optical filter configuration. Collect one image
303
More than one image per optical configuration?
Yes
304
Move to new position within well and collect another image
Yes
305
any more images per well to collect?
No
306
Return stage to starting position within well, montage all collected images, give unique file name, save and close file
307
any more optical channels to collect in this well?
Yes
308
Select next optical filter configuration. Collect one image
No
309
Any more wells to visit?
Yes
310
Move to next well
No
311
Any more plates ?
No
312
Backup data onto CD rom or other media
Stop
Yes

Fig. 7J

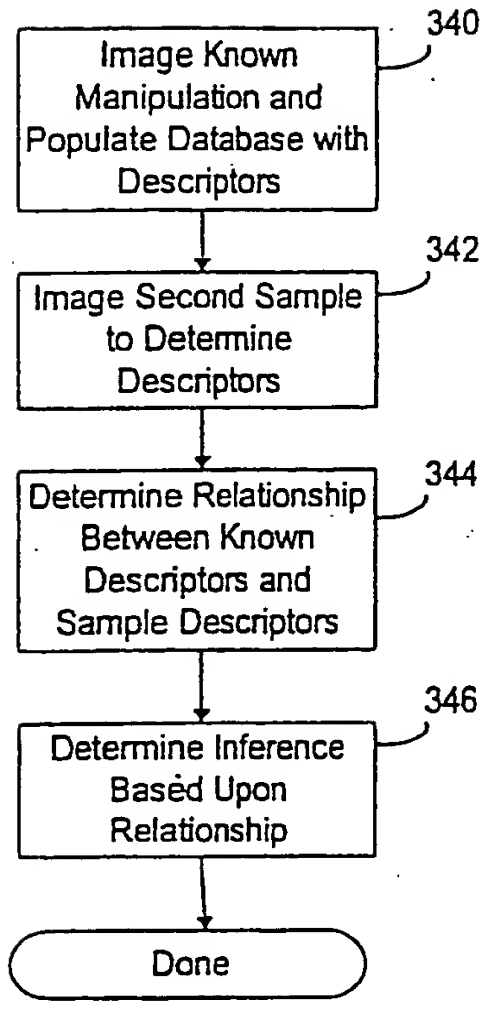


Image Known Manipulation and Populate Database with Descriptors
340
Image Second Sample to Determine Descriptors
342
Determine Relationship Between Known Descriptors and Sample Descriptors
344
Determine Inference Based Upon Relationship
346
Done

Fig. 7K

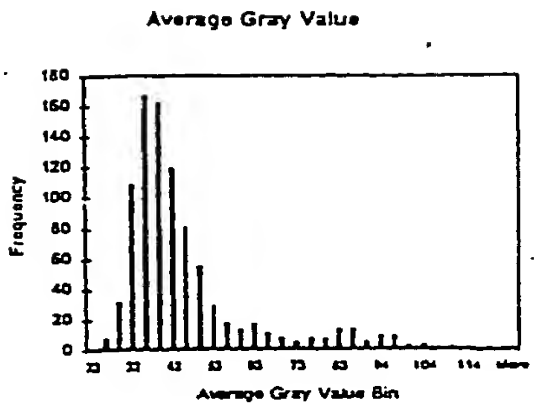


Fig. 8A

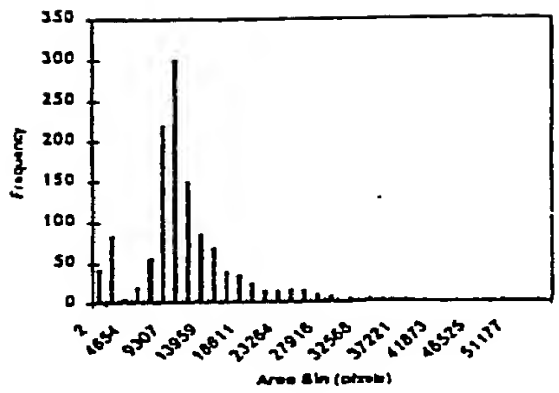


Fig. 8B

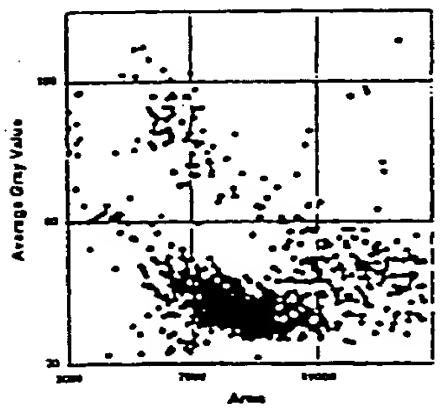


Fig. 8C

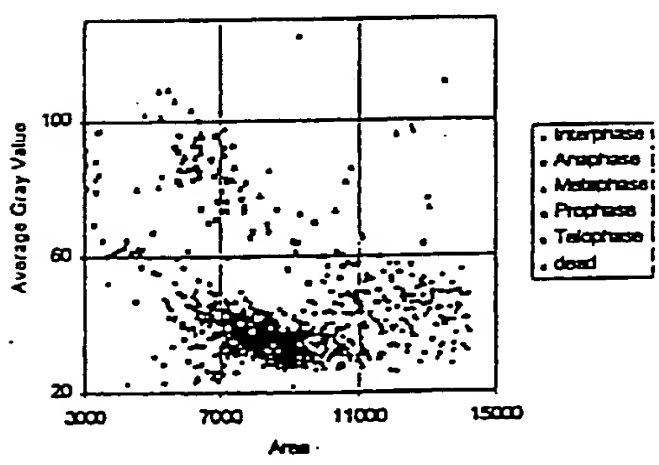


Fig. 8D

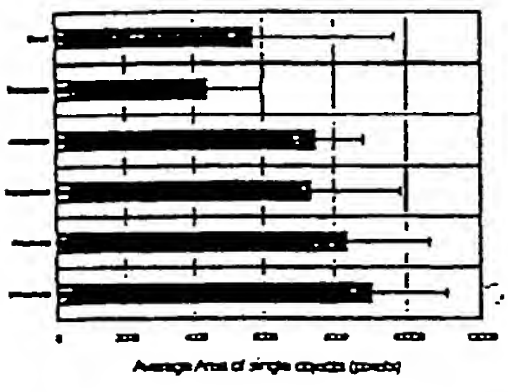


Fig. 8E

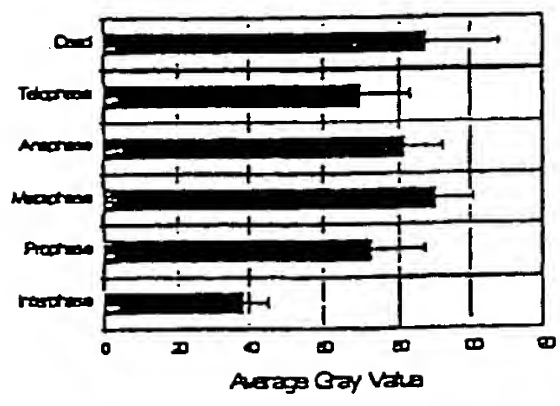


Fig. 8F

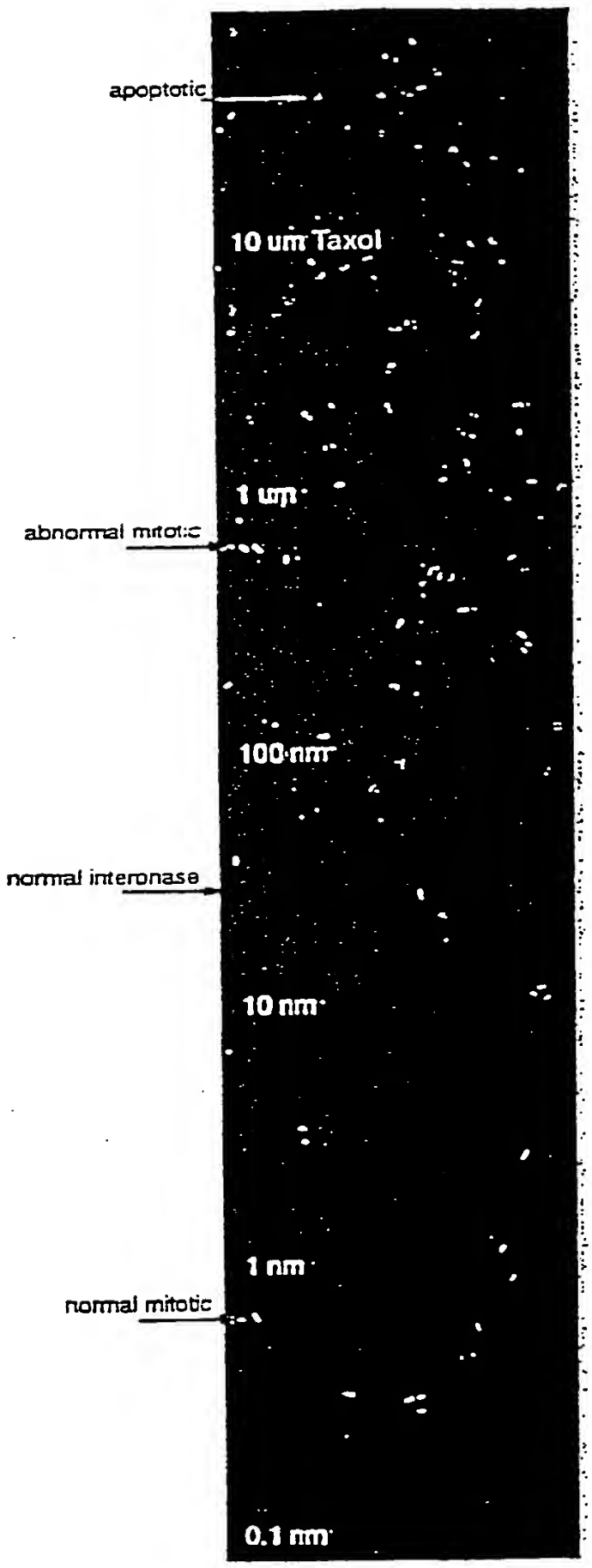

apoptotic
10 um Taxol
1 um
abnormal mitotic
100 nm
normal interonase
10 nm
1 nm
normal mitotic
0.1 nm

Fig. 9

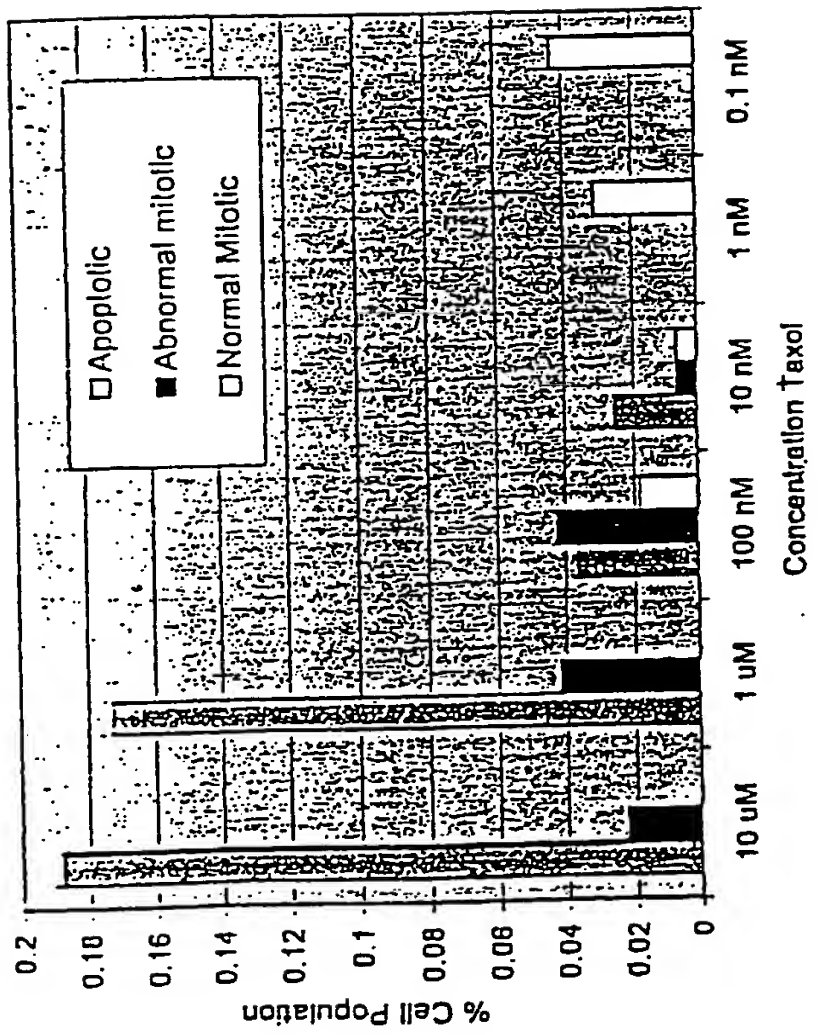

MDCK cells treated with Taxol for 4.5 hours
Apoptotic
Abnormal mitotic
Normal Mitotic
% Cell Population
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
10 uM
1 uM
100 nM
10 nM
1 nM
0.1 nM
Concentration Taxol

Fig. 10

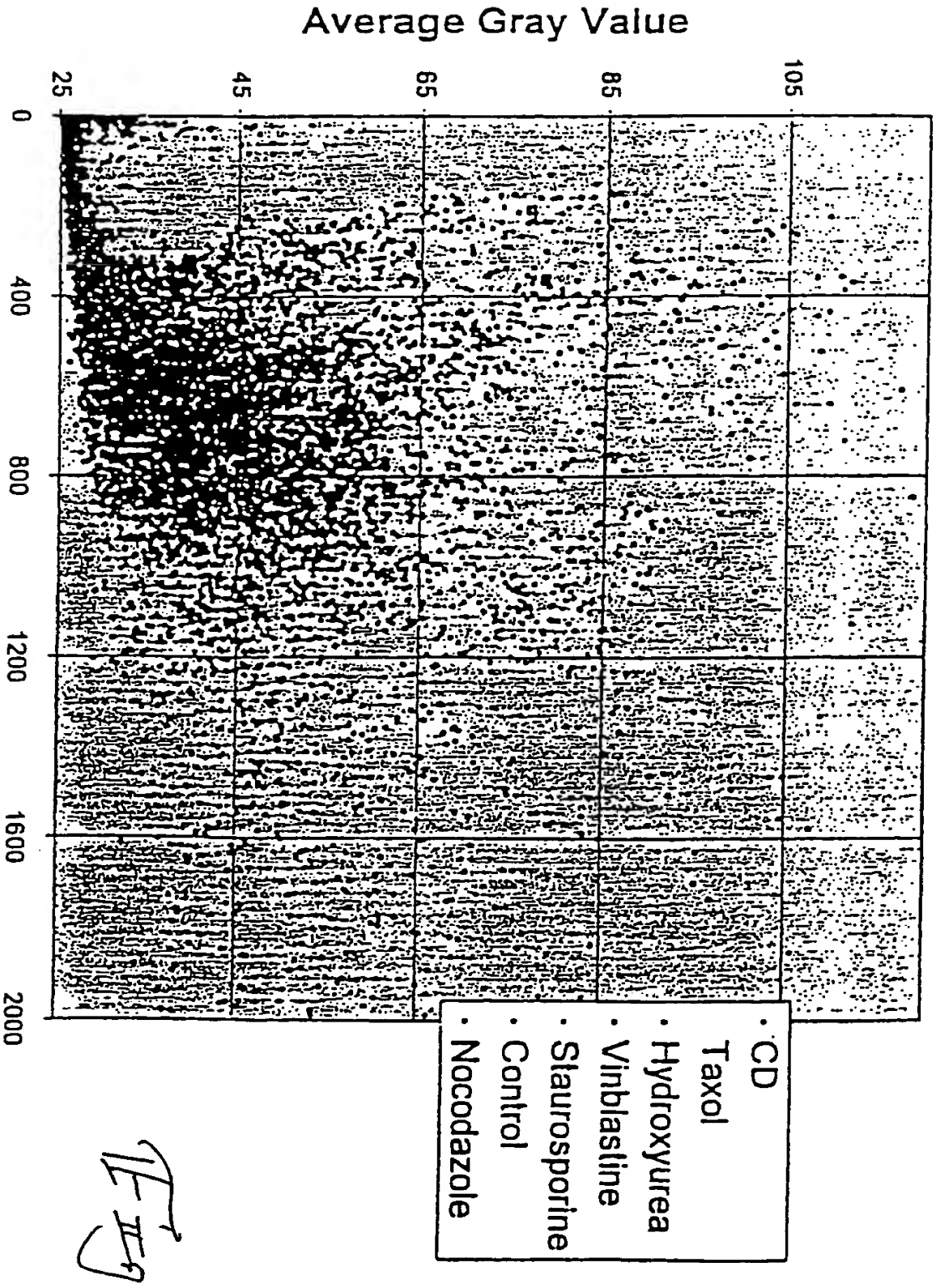

Average Gray Value
25
45
65
85
105
0
400
800
1200
1600
2000
Area
Scatter Plot
CD
Taxol
Hydroxyurea
Vinblastine
Staurosporine
Control
Nocodazole

Fig 11

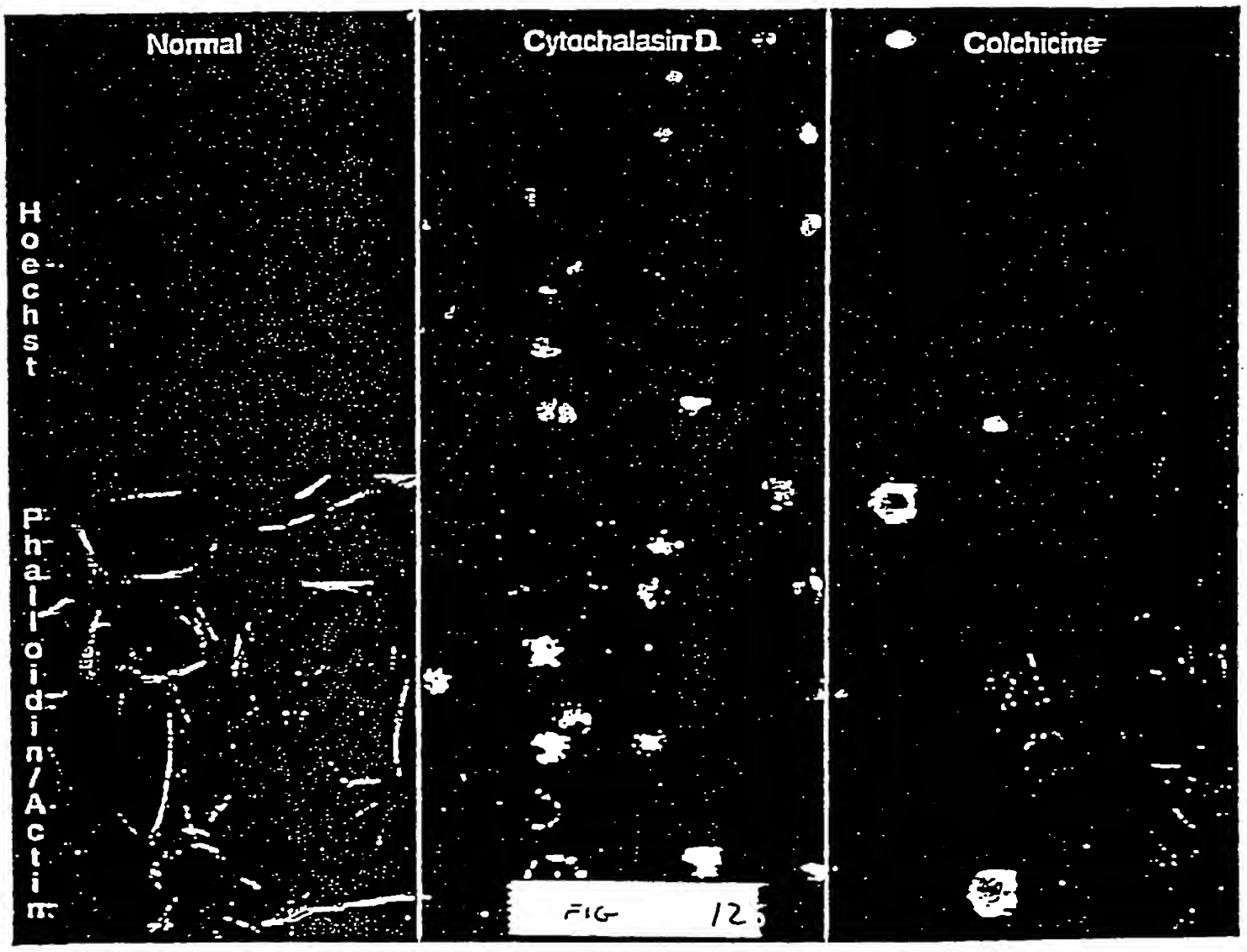

Normal
Cytochalasin D
Colchicine
Hoechst
Phalloidin/Actin

FIG 12

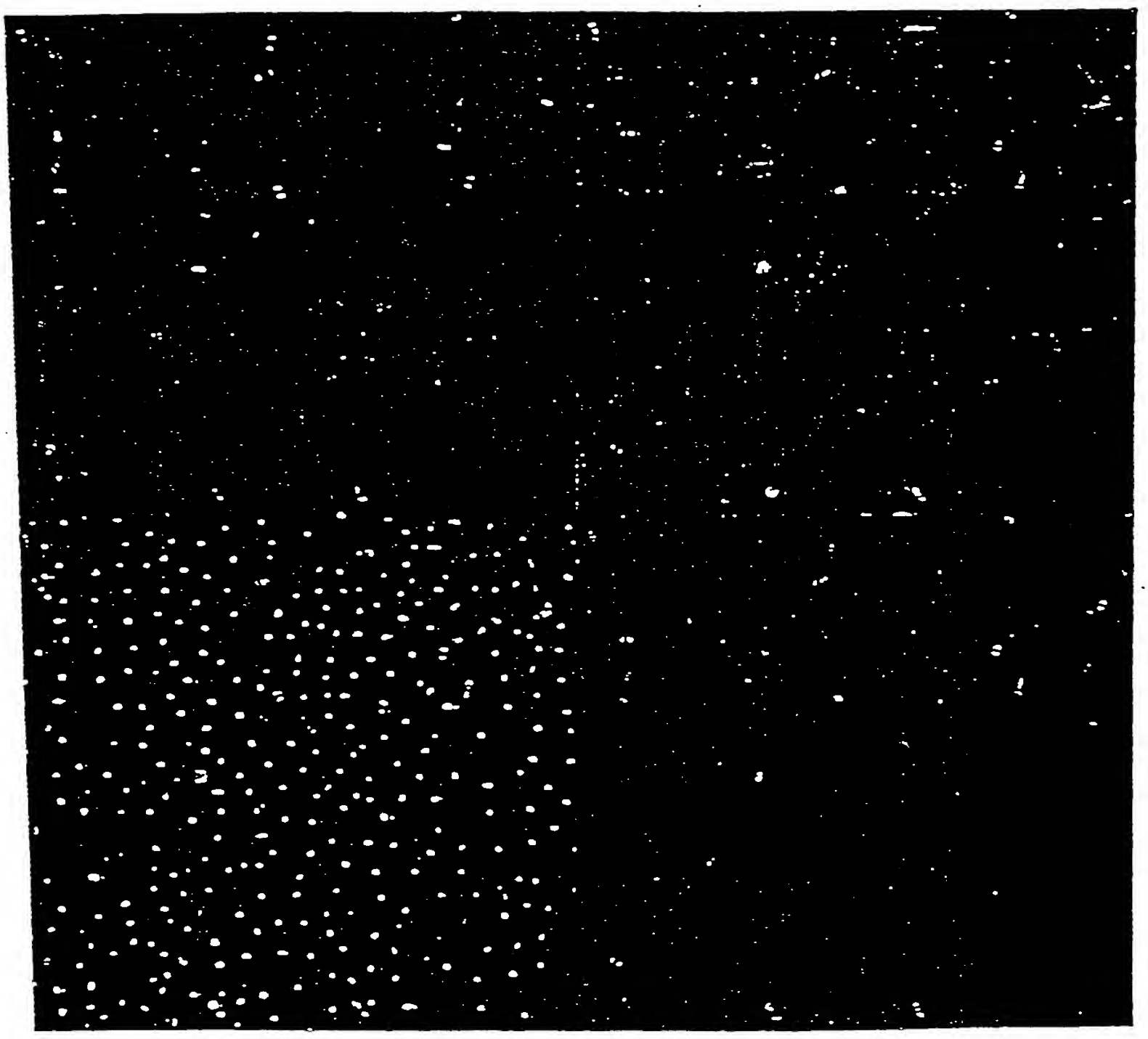


Fig 13

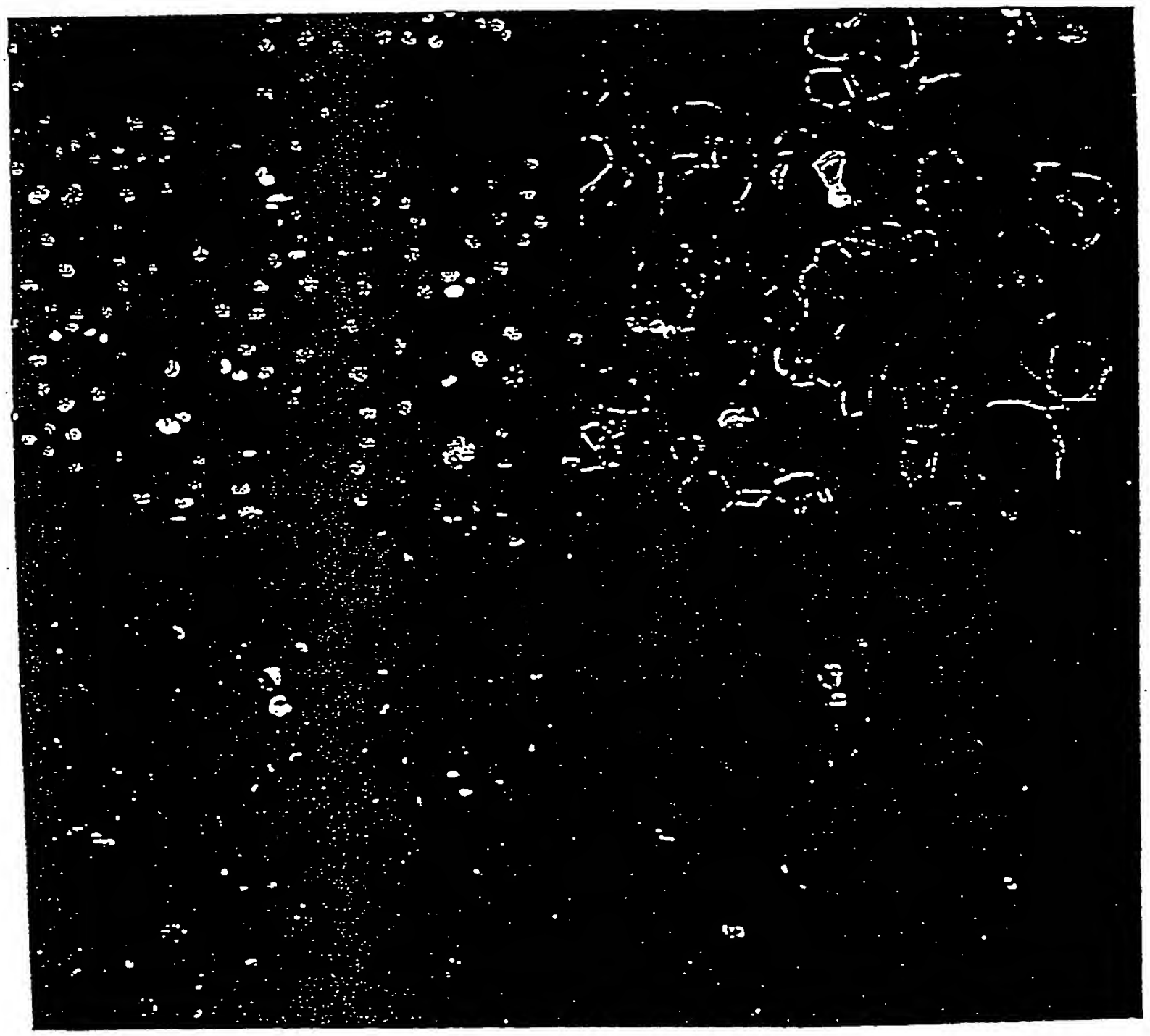


Fig 14

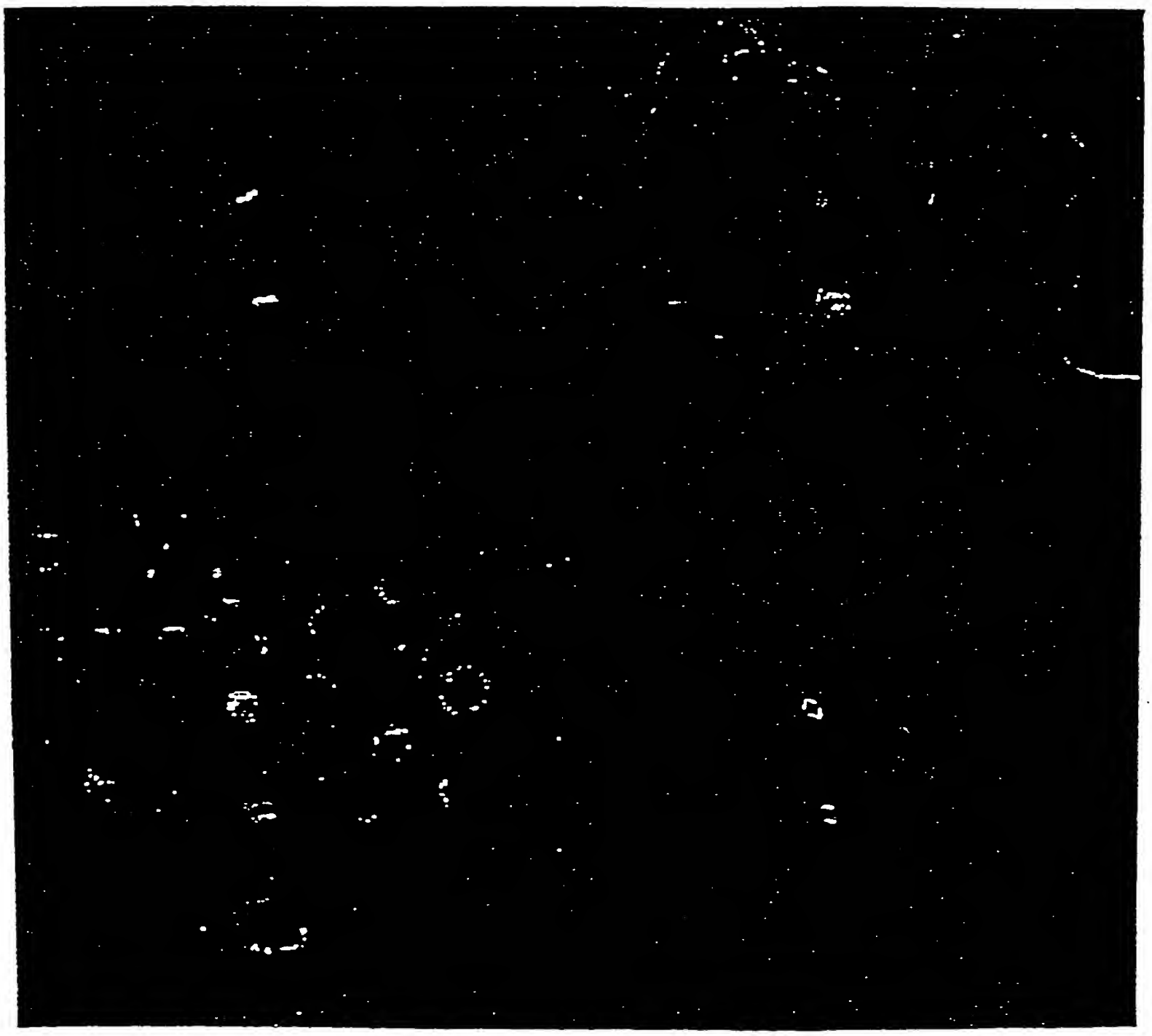


Fig 15

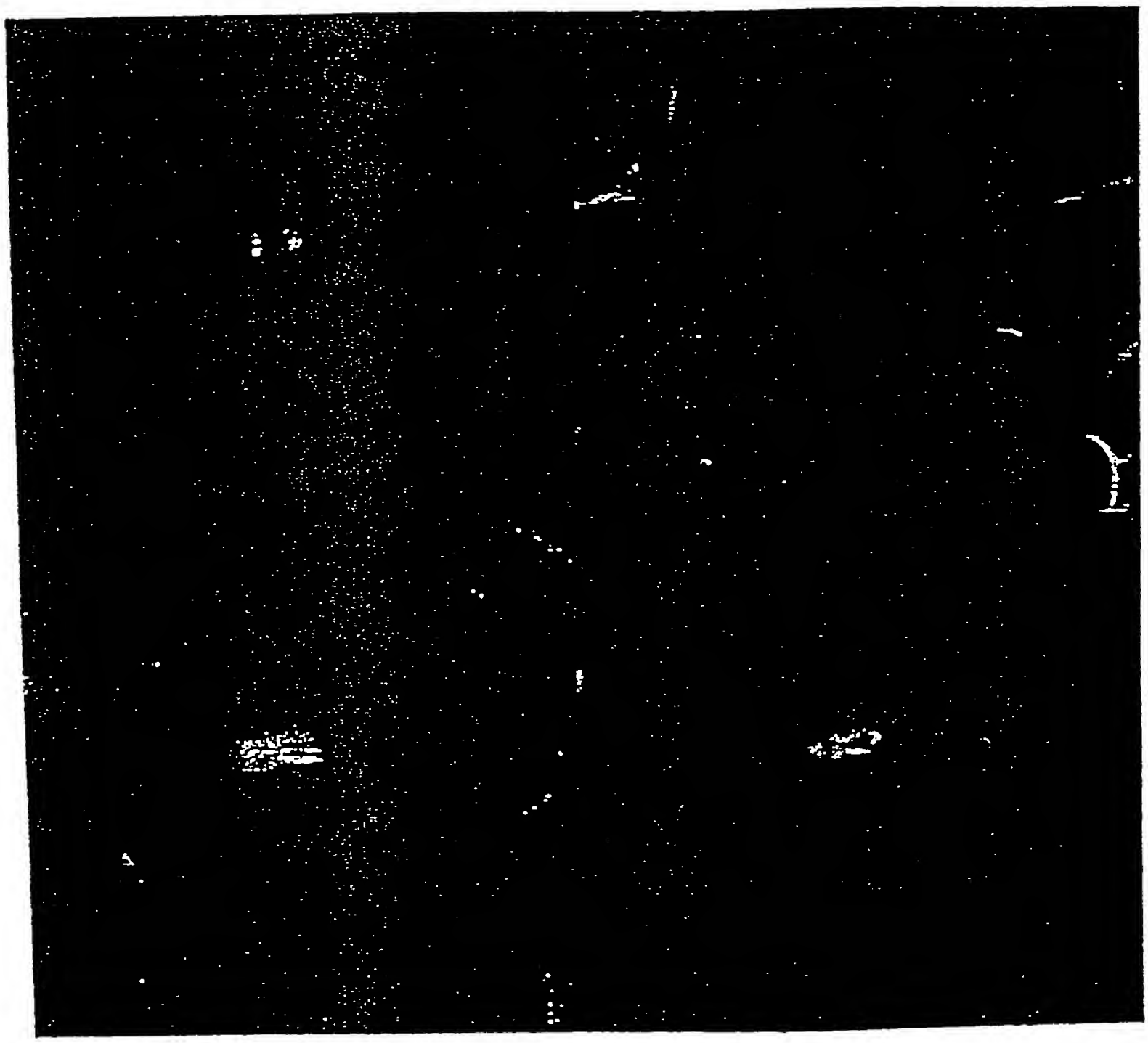


Fig 16

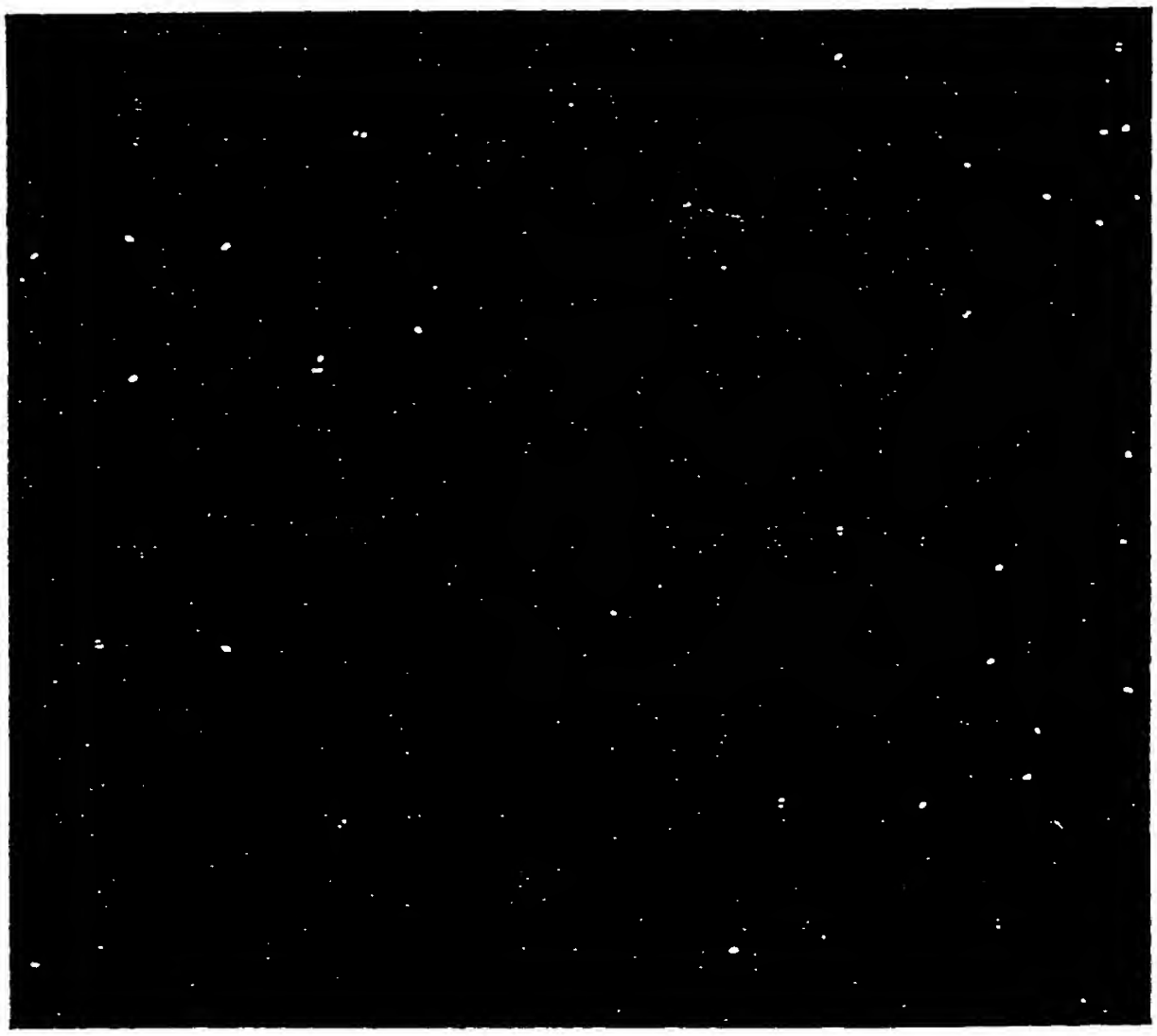


Fig 17

Conversion of morphometric parameters into nucleic acid code and clustering of the resulting sequences using Neighbor Joining method.

Compound:	Measurements																									
	Count	Area	Perimeter	Length	Breadth	Fiber length	Fiber breadth	Shape factor	Ell. form factor	Inner radius	Outer radius	Mean radius	Equiv. radius	Equiv. sphere vol.	Equiv. prolate vol.	Equiv. oblate vol.	Equiv. sphere surface area	Average gray value	Total gray value	Optical density	Radial dispersion	Texture Difference Moment	EFA Harmonic 2, Semi-Maj	EFA Harmonic 2, Semi-Min
Control	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	a	t	t
Taxol	a	t	t	t	t	t	t	t	a	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t
CD	c	a	a	a	t	a	t	t	c	a	a	a	a	a	a	a	a	t	a	a	a	t	a	g
Nocodazol	c	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t
Staurosponne	g	g	c	a	a	t	a	a	t	g	a	a	a	t	g	g	g	a	a	t	a	t	a	a
Vinblastine	c	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	t	g	t	t	t	t	t
Hydroxyurea	g	t	t	t	t	t	t	g	t	t	t	t	t	t	t	t	t	t	t	c	t	a	t	t

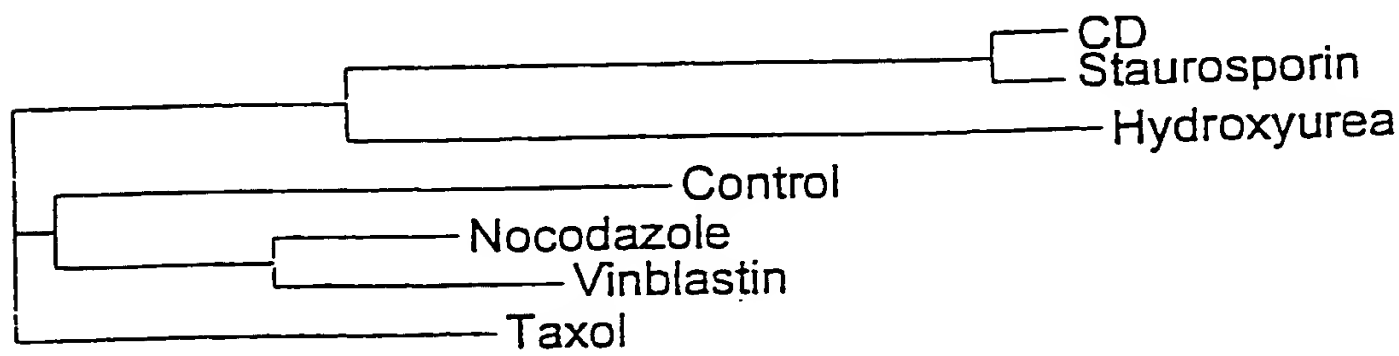


Fig 10

Conversion of morphometric parameters into amino acid codes and clustering of the resulting sequences using Neighbor Joining method.

	Count	Area	Perimeter	Length	Breadth	Fiber length	Fiber breadth	Shape factor	Ell. form factor	Inner radius	Outer radius	Mean radius	Equiv. radius	Equiv. sphere vol.	Equiv. prolate vol.	Equiv. oblate vol.	Equiv. sphere surface a	Average gray value	Total gray value	Optical density	Radial dispersion	Texture Difference Mo	EFA Harmonic 2, Seml-	EFA Harmonic 2, Seml-	
control	I	P	T	T	N	S	D	W	E	S	T	T	T	F	C	C	P	R	M	C	T	G	T	T	Y
Taxol	G	F	M	M	P	M	P	H	G	S	M	M	W	C	F	P	F	R	C	M	M	H	M	P	S
CD	F	G	G	G	M	G	M	K	A	G	G	G	G	G	G	G	G	H	G	G	G	M	G	V	H
Nocodazol	W	F	M	M	W	M	P	T	R	S	M	M	M	F	M	W	F	M	M	R	M	M	M	F	G
Staurosporine	N	V	A	G	G	M	G	G	Y	V	G	G	G	M	V	V	V	G	G	H	G	M	G	G	V
Vinblastine	F	W	W	M	W	W	C	W	D	S	M	W	W	M	M	M	W	M	V	E	M	M	M	F	P
Hydroxyurea	S	H	H	H	H	H	H	V	H	H	H	H	H	H	H	H	H	H	H	A	H	G	H	H	D

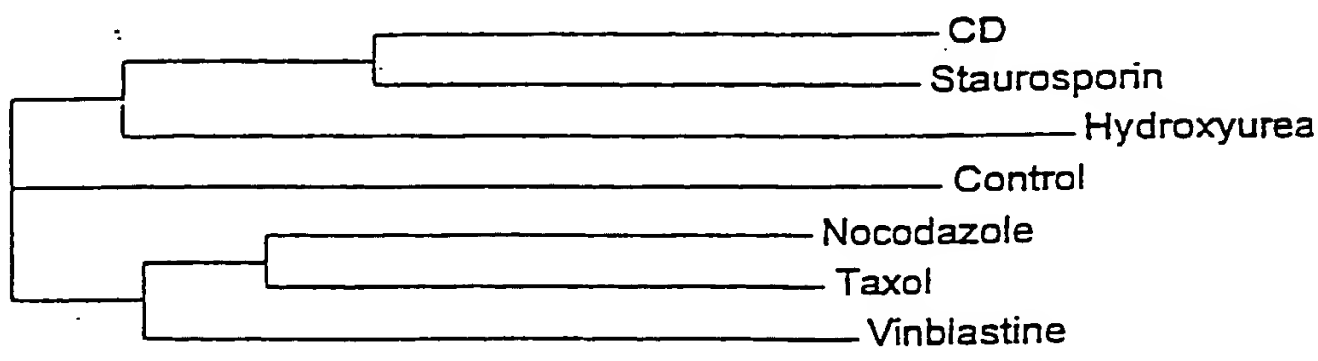


Fig 19

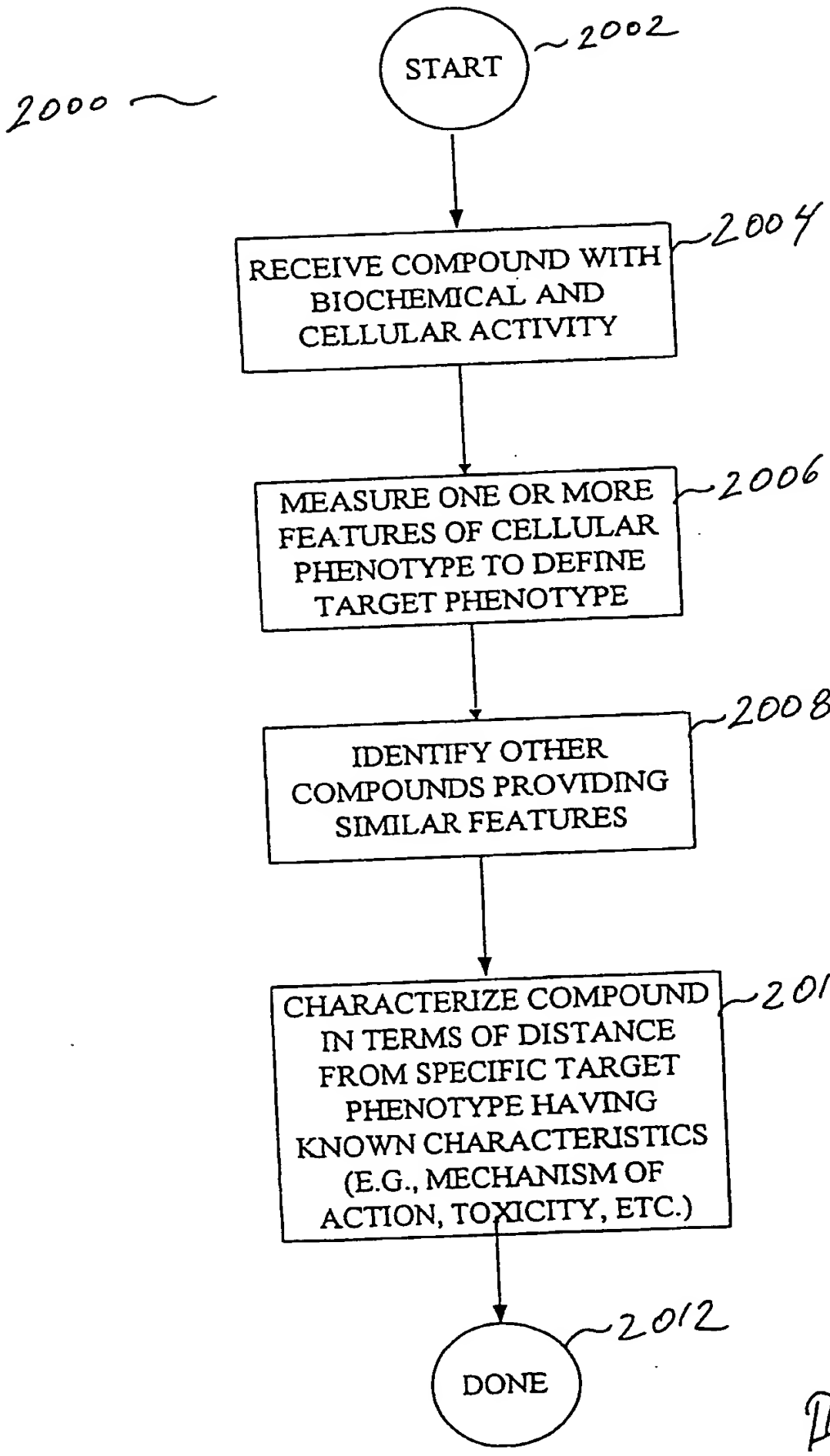

2000
START
2002
RECEIVE COMPOUND WITH BIOCHEMICAL AND CELLULAR ACTIVITY
2004
MEASURE ONE OR MORE FEATURES OF CELLULAR PHENOTYPE TO DEFINE TARGET PHENOTYPE
2006
IDENTIFY OTHER COMPOUNDS PROVIDING SIMILAR FEATURES
2008
CHARACTERIZE COMPOUND IN TERMS OF DISTANCE FROM SPECIFIC TARGET PHENOTYPE HAVING KNOWN CHARACTERISTICS (E.G., MECHANISM OF ACTION, TOXICITY, ETC.)
2010
DONE
2012

Fig 20

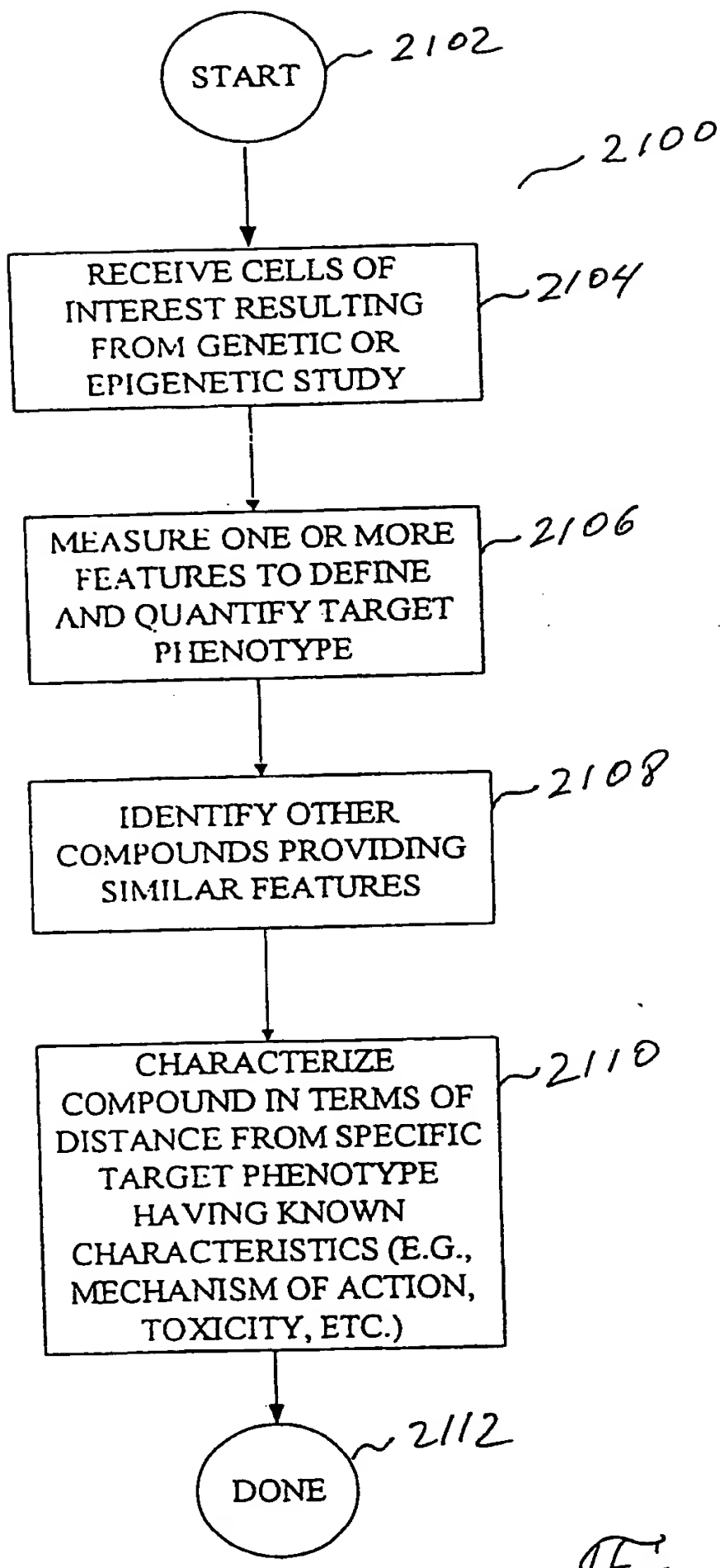

START
2102
2100
RECEIVE CELLS OF INTEREST RESULTING FROM GENETIC OR EPIGENETIC STUDY
2104
MEASURE ONE OR MORE FEATURES TO DEFINE AND QUANTIFY TARGET PHENOTYPE
2106
IDENTIFY OTHER COMPOUNDS PROVIDING SIMILAR FEATURES
2108
CHARACTERIZE COMPOUND IN TERMS OF DISTANCE FROM SPECIFIC TARGET PHENOTYPE HAVING KNOWN CHARACTERISTICS (E.G., MECHANISM OF ACTION, TOXICITY, ETC.)
2110
DONE
2112

Fig 21

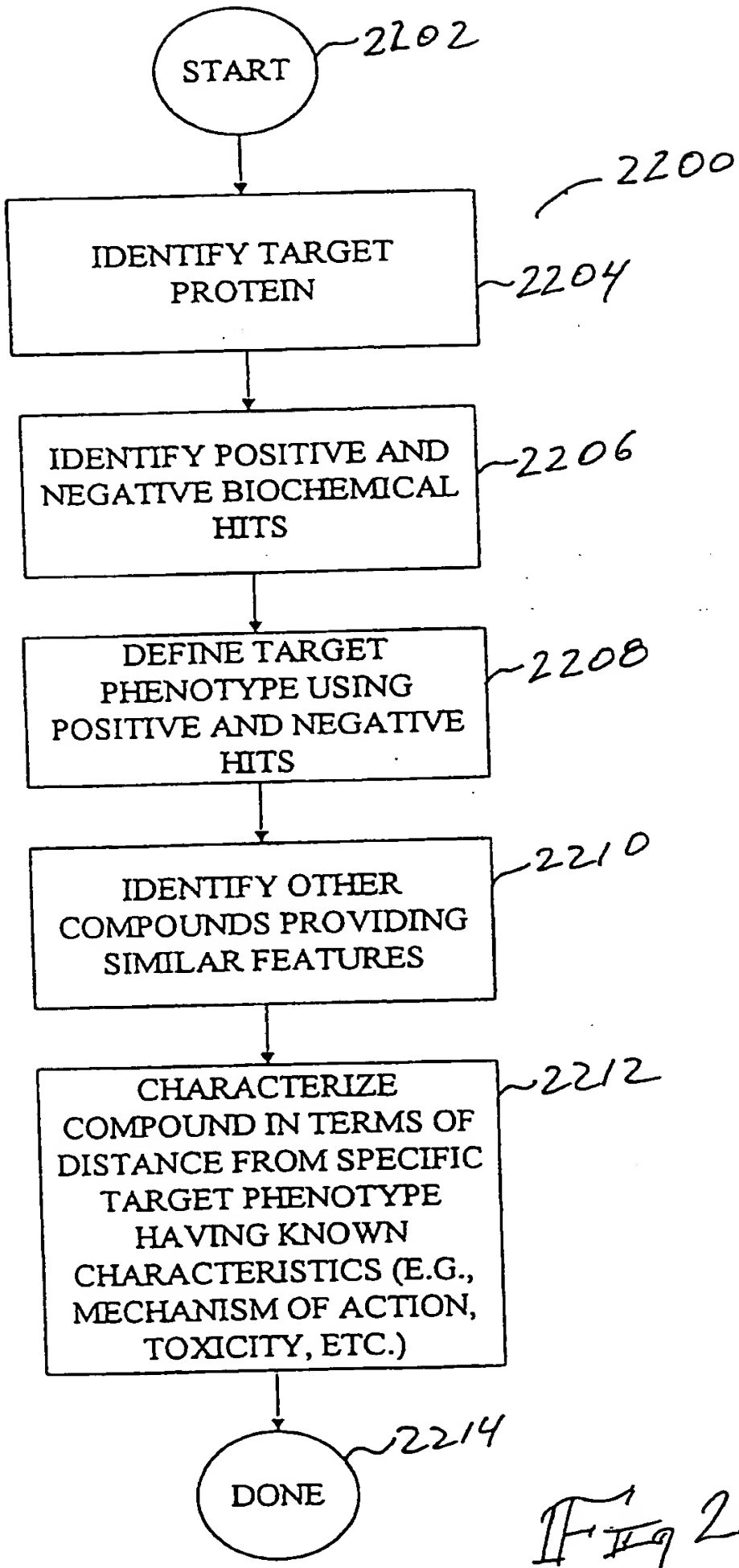

START
2202
2200
IDENTIFY TARGET PROTEIN
2204
IDENTIFY POSITIVE AND NEGATIVE BIOCHEMICAL HITS
2206
DEFINE TARGET PHENOTYPE USING POSITIVE AND NEGATIVE HITS
2208
IDENTIFY OTHER COMPOUNDS PROVIDING SIMILAR FEATURES
2210
CHARACTERIZE COMPOUND IN TERMS OF DISTANCE FROM SPECIFIC TARGET PHENOTYPE HAVING KNOWN CHARACTERISTICS (E.G., MECHANISM OF ACTION, TOXICITY, ETC.)
2212
DONE
2214

Fig 22